

OM of: US-09-652-292-2 to: GenEmbl.* out_format : pfs

Date: Mar 15, 2002 7:52 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters: -dbv-xlh

-MODEL=framet_p2n.model -DBV=xlh
-Q/cgnt2/_USPTO_spool/US09652292/runat_13032002_161725_3207/app_query.fasta_1.606
-MINMATCH=1 -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-DBNAME=GenB-0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000
-GAPEXT=7.000 -GAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=bloum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09652292 -CGN1_1.6252
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-652-292-2

Query length: 541

Database: GenEmbl.*

Database sequences: 1472140

Database length: -341344837

Search time (sec): 1293.340000

score_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
gb_pr:HS28H201	+ 2765.00	2545.58	1.6e-133	4126	Novel human gene map
gb_pr:AF231240	+ 2765.00	2545.50	1.6e-133	4167	AF231240 Homo sapiens facilit
gb_pr:AF248053	+ 2765.00	2545.09	1.7e-133	4396	AF248053 Homo sapiens glucose
gb_pr:HS28H20	+ 2286.50	2077.89	1.8e-107	127418	AL031055 Human DNA sequence
gb_hg:AL591064	+ 1678.00	1513.69	4.9e-76	200125	AL591064 Mus musculus chromo
gb_hg:AL591805	+ 1678.00	1513.17	5.9e-76	243075	AL591805 Mus musculus chromo
gb_pr:AB056798	+ 932.00	855.94	2.1e-39	4521	AB056798 Macaca fascicularis b
gb_pr:AL449363	+ 727.00	640.15	2.2e-27	141289	AL449363 Human DNA sequence
gb_hg:AC023293	+ 724.00	636.25	3.7e-27	163569	AC023293 Homo sapiens clone
gb_ba:BS292954	+ 626.50	569.81	1.8e-23	18189	Z99254 B. subtilis yws(A,B,C,D)
gb_ba:BSUB0019	+ 626.50	544.36	4.8e-22	212610	Z99122 Bacillus subtilis com
gb_ba:BSUB0021	+ 595.00	515.23	2.0e-20	215534	Z99124 Bacillus subtilis com
gb_ba:AB005554	+ 592.00	526.35	4.8e-21	36448	AB005554 Lactococcus lactis g
gb_ba:AF045552	+ 571.00	521.16	9.4e-21	5957	AF045552 Lactobacillus brevis
gb_pat:AX076673	+ 556.00	515.20	2.0e-20	2177	AX076673 Sequence 7 from Paten
gb_pr:HSN315644	+ 556.00	512.04	3.0e-20	3261	AJ315644 Homo sapiens mRNA for
gb_pat:AX076675	+ 547.00	505.82	6.7e-20	2504	AX076675 Sequence 9 from Paten
gb_ro:RNO315643	+ 547.00	505.82	6.7e-20	2504	AJ315643 Rattus norvegicus mRN
gb_ba:AE006381	+ 541.50	488.47	6.2e-19	12049	AE006381 Lactococcus lactis s
gb_pl:AF280431	+ 529.00	489.49	5.5e-19	2422	AF280431 Mesembryanthemum crys
gb_pl:AF280432	+ 526.50	488.63	6.1e-19	2013	AF280432 Mesembryanthemum crys
gb_pl:SPTR2	+ 520.50	481.91	1.4e-18	3244	X99105 S. pombe ITR2 gene. 7/1
gb_pl:SPAC2068	+ 520.50	461.94	1.9e-17	30200	Z95334 S. pombe chromosome 1 c
gb_pl:AF215852	+ 519.00	482.62	1.3e-18	1795	AF215852 Nicotiana tabacum hex
gb_pat:AX172681	+ 517.50	480.77	1.7e-18	1905	AX172681 Sequence 171 from Pat
gb_pl:AF215853	+ 516.00	480.51	1.7e-18	1650	AF215853 Solanum tuberosum hex
gb_pat:AX008642	+ 516.00	480.50	1.7e-18	1653	AX008642 Sequence 3 from Paten
gb_pl:AF280906	+ 515.00	477.93	2.4e-18	2041	AF280906 Mesembryanthemum crys
gb_hg:AC093093	+ 509.00	440.07	3.1e-16	127914	AC093093 Oryza sativa chromo
gb_pl:AC083945	+ 509.00	438.94	3.6e-16	147706	AC083945 Oryza sativa strain
gb_pat:AX008644	+ 507.00	471.27	5.7e-18	1864	AX008644 Sequence 5 from Paten
gb_pl:AF215851	+ 507.00	471.27	5.7e-18	1864	AF215851 Spinacia oleracea hex
gb_pat:AX120287	+ 502.00	468.50	8.1e-18	1473	AX120287 Sequence 203 from Pat
gb_pat:AX120085	+ 502.00	425.75	1.9e-15	349980	AX120085 Sequence 1 from Pat
gb_pl:ATC19444	+ 501.50	429.71	1.2e-15	198735	AJ161544 Arabidopsis thalian
gb_pl:ATFCA6	+ 501.50	449.14	1.3e-15	213788	Z97341 Arabidopsis thaliana
gb_ba:ECU00377	+ 499.00	449.12	9.7e-17	12354	AE000377 Escherichia coli K12
gb_ba:ECU28377	+ 499.00	430.05	1.1e-15	141744	I28377 Escherichia coli K-12
gb_in:LETD17RA	+ 497.50	439.34	2.6e-17	2798	M85072 Leishmania donovani D1
gb_pl:AC006068	+ 497.50	432.27	8.4e-16	89338	AC006068 Arabidopsis thaliana

gb_ba:AE005524 + 496.00 447.13 1.3e-16 11178 AE005524 Escherichia coli
gb_ba:AF002563 + 496.00 422.35 3.0e-15 266658 AF002563 Escherichia coli
gb_ba:ECOARAE4 + 495.00 456.85 3.6e-17 2866 J03732 E.coli arabinose-pro
gb_ba:AE000368 - 495.00 447.24 1.2e-16 9800 AE000368 Escherichia coli K
gb_ba:AE005513 - 495.00 446.51 1.4e-16 10755 AE005513 Escherichia coli
seq_name: gb_pr:HS28H201
seq_documentation_block:
LOCUS HS28H201 4126 bp mRNA PRI 21-NOV-2000
DEFINITION Novel human gene mapping to chromosome 20, similar to membrane
transporters.
ACCESSION AL137188
VERSION AL137188.3 GI:11322734
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 4126)
AUTHORS Stavrides,G.S., Hashim,Y., Huckle,E.J. and Deloukas,P.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2000) E-mail contact: humquery@sanger.ac.uk
COMMENT On Nov 23, 2000 this sequence version replaced gi:11065679.
This cDNA sequence was assembled from public domain ESTs and single
pass sequencing reads from expressed DNA templates, aligned to the
genomic DNA sequence from the bacterial clone 28H20 (AL031055). The
EST sequences listed match this sequence with an identity of at
least 95% between the coordinates shown.
Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20/ Sanger Centre name :
dJ28H20.C20.1.

FEATURES
source 1. 4126
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q"
1. 112
/number=1
9. 1634
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB69822.2"
/db_xref="GI:11065680"
/translation="MGHSPVPLCASVSLGGTFFGYELAVISGALLPLQLDPLGLSC
LEQFLVSGSLIALLASLVGGFLIDCVGRKQAILGNLVLGLAGSLTGLAGSLAWLV
LGRAVGFATISLSMACCIYVSELVGRQGVLSVYEAGITVGGSEAPKLGPRYSF
PWGRHFGNATAPVLSQSLLELFLPAGTDTATKDLIPLOGSEAPKLGPRYSF
LDLFRADNHRRTVGLGLVLFPOOLTQOPNVLCTASTFSSVFGHSGSVAVASVGL
GAVKAAATLTAMGLVDRAIRALLAGCALMALSVGILGVSAFVPMDSGPPSLAVPN
ATGQTGLPGSGLIQDSLSLP IPTNEDQREPISTAKTKPHRSGDPSAPRIALS
SALCPPLPARGHALLRWLTALLCLMVFVSFGFPGVTLVLSVPEIRGRAFAF
CNSPNAANLPISLFDLIGTIGLSMTFLYGLTAVLGLGFVILFVPETKGQSLAEI
DQPKRRFTLSFGHRSNGTGPISRIEISAS"
13. 1296
/number=2
1297. 1419
/number=3
1338. 1643
/note="matches EST AA313045"
1420. 1555
/number=4
1556. 4126
/number=5
2095. 2137
/note="matches EST AA714011 from clone IMAGE:1240821"
complement(2142..2172)
/note="matches EST BE747194 from clone IMAGE:3929520"
join(2274..2544,2548..2589,2646..2763)
/note="matches EST AA007343 from clone IMAGE:429263"
complement(2311..2352)
/note="matches EST AV650272 from clone GLCCDA12

```
misc_feature matches EST AV650673 from clone GLCCHC08
misc_feature matches EST AV650603 from clone GLCCGEI0
misc_feature matches EST AV650406 from clone GLCCGE02
misc_feature matches EST AV650686 from clone GLCCHD09
misc_feature join(2417..2711..2814..3001..3029)
misc_feature /note="matches EST W02942 from clone IMAGE:291802"
misc_feature /note="matches EST W31922 from clone IMAGE:320524"
misc_feature join(2476..2695..2693..2817)
misc_feature /note="matches EST AA489718 from clone IMAGE:823660"
misc_feature join(2647..2694..2693..3059..3053..3115)
misc_feature /note="matches EST AA403072 from clone IMAGE:758347"
misc_feature join(2739..3054..3050..3143)
misc_feature /note="matches EST AA232787 from clone IMAGE:666656"
misc_feature join(2757..2884..2881..3023)
misc_feature /note="matches EST N40351 from clone IMAGE:269966"
misc_feature complement(join(2794..2876..2894..2935..2935..3045..3046..3138..3135..3249))
misc_feature /note="matches EST AA007344 from clone IMAGE:429263"
misc_feature complement(2796..3249)
misc_feature /note="matches EST A1334230 from clone IMAGE:1932055"
misc_feature 2808..2999
misc_feature /note="matches EST N44116 from clone IMAGE:272953"
misc_feature complement(join(2810..3045..3046..3250))
misc_feature /note="matches EST A1082631 from clone IMAGE:1660573"
misc_feature complement(join(2810..2885..2911..2935..3091..3250))
misc_feature /note="matches EST N27535 from clone IMAGE:269966"
misc_feature complement(join(2818..3045..3046..3092..3091..3249))
misc_feature /note="matches EST N36110 from clone IMAGE:272953"
misc_feature complement(join(2905..3045..3046..3138..3135..3249))
misc_feature /note="matches EST AA489618 from clone IMAGE:823660"
misc_feature complement(join(2918..3045..3046..3138..3135..3250))
misc_feature /note="matches EST AA233364 from clone IMAGE:666656"
misc_feature complement(3040..3549)
misc_feature /note="matches EST AW298226 from clone IMAGE:2733069"
misc_feature join(3212..3478..3540..3563)
misc_feature /note="matches EST W38959 from clone IMAGE:304871"
misc_feature complement(3397..3974)
misc_feature /note="matches EST AA628914 from clone IMAGE:1032940"
misc_feature 3402..3738
misc_feature /note="matches EST C04258 from clone 3NHC3019"
misc_feature join(3459..3633..3633..3656)
misc_feature /note="matches EST AA045035 from clone IMAGE:488780"
misc_feature 3459..3877
misc_feature /note="matches EST AAL34031 from clone IMAGE:503848"
misc_feature join(3542..3710..3754..3948)
misc_feature /note="matches EST W39026 from clone IMAGE:305523"
misc_feature complement(3576..4126)
misc_feature /note="matches EST A1042706 from clone IMAGE:1431595"
misc_feature join(3596..3809..3798..4045)
misc_feature /note="matches EST AAL33497 from clone IMAGE:490637"
misc_feature complement(join(3596..3619..3642..4038))
misc_feature /note="matches EST AAL115737 from clone IMAGE:490637"
misc_feature complement(3597..4126)
misc_feature /note="matches EST A1041537 from clone IMAGE:1643803"
misc_feature complement(3608..3974)
misc_feature /note="matches EST A1088144 from clone IMAGE:1683131"
misc_feature complement(3625..4126)
misc_feature /note="matches EST AA404352 from clone IMAGE:758347"
misc_feature complement(join(3626..3645..3674..3898..3867..4126))
misc_feature /note="matches EST AAL33966 from clone IMAGE:503848"
misc_feature complement(join(3642..3898..3877..4122))
misc_feature /note="matches EST AW973035"
misc_feature complement(join(3685..3898..3867..4126))
misc_feature /note="matches EST A1081145 from clone IMAGE:1680293"
misc_feature complement(join(3689..3898..3867..4126))
misc_feature /note="matches EST A1753293 from clone HBMSC_cr08D03"
misc_feature complement(join(3694..3898..3867..4123))
misc_feature /note="matches EST A1097288 from clone IMAGE:1707216"
misc_feature complement(3727..4126)
misc_feature /note="matches EST A1292321 from clone IMAGE:1894742"
misc_feature complement(join(3737..3898..3867..4126))
misc_feature /note="matches EST A1753932 from clone HBMSC_cr16C08"
```

```
misc_feature complement(join(3774..3898..3867..4126))
misc_feature /note="matches EST A1277131 from clone IMAGE:1893578"
misc_feature complement(3793..4126)
misc_feature /note="matches EST N93207 from clone IMAGE:304871"
misc_feature complement(3795..4124)
misc_feature /note="matches EST A1289525 from clone IMAGE:1992441"
misc_feature complement(join(3795..3887..3877..4017..4005..4126))
misc_feature /note="matches EST AW103571 from clone IMAGE:2614034"
misc_feature complement(3867..4125)
misc_feature /note="matches EST AW166863 from clone IMAGE:2634465"
misc_feature complement(3867..4123)
misc_feature /note="matches EST AA532534 from clone IMAGE:986291"
misc_feature complement(3867..4126)
misc_feature /note="matches EST A1753418 from clone HBMSC_cr10B04"
misc_feature complement(3945..4126)
misc_feature /note="matches EST A1240819 from clone IMAGE:1848499"
BASE COUNT 914 a 1066 c 979 g 1167 t
ORIGIN
alignment_scores:
    Quality: 2765.00      Length: 541
    Ratio: 5.111          Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-09-652-292-2 x HS28H201 ..
Align seg 1/1 to: HS28H201 from: 1 to: 4126
1 MetGlyHisSerProProValLeuProLeuCysAlaSerValSerLeuLeu 17
|||||
9 ATGGGCACTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 58
|||||
17 uGlyGlyLeuThrPheClyTyrGluLeuAlaValIleSerGlyAlaLeuL 34
|||||
59 GGCTGGCCTGACCTTTGGTTATCAACTGGCAGTCATATCAGGTGCCTGC 108
|||||
34 euProLeuGlnLeuAspPheGlyLeuSerCysLeuGluGlnGluPheLeu 50
|||||
109 TGCACATGACGCTGACTTTGGGCTAAGCTGCTTGGAGCAGGAGTCCCTG 158
|||||
51 ValGlySerLeuLeuGlyAlaLeuLeuAlaLeuValGlyGlyPh 67
|||||
159 GTGGCAGCCTGCTCTCTGGGGCTCTCTCTGCTGCTGCTGCTGCTGCT 208
|||||
67 eLeuIleAspCysTyrGlyArgIysGlnAlaLeuGlySerAsnLeuV 84
|||||
209 CCTCATPTGACTGCTATGGCAGGAAGCAAGCCATCTCGGGAGCACTTGG 258
|||||
84 aLeuLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAlaTrp 100
|||||
259 TGCTGCTGGCAGCAGCCTGACCTGGGCTGCTGGTTCCTGCTGCTGCTG 308
|||||
101 LeuValLeuGlyArgAlaValValGlyPheAlaIleSerLeuSerSerMe 117
|||||
309 CTGGTCTCTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
|||||
117 tAlaCysCysIleTyrValSerGluLeuValGlyProArgGlnArgGlyV 134
|||||
359 GCGTTCGCTGTATCTACGTGTACAGCTGCTGGGGCCACGCGACGGGAG 408
|||||
134 aLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeuSer 150
|||||
409 TGCTGGTGTCCCTCTATGAGGCGAGCATCACCCTGGGCTATCCTGCTCC 458
|||||
151 TyrAlaLeuAsnTyrAlaLeuAlaGlyThrProTyrGlyTyrArgHisMe 167
|||||
459 TATGCCCTCAACTATGCACTGGCTGGTACCCCTCGGGATGGAGGCACAT 508
|||||
167 tPheGlyTyrAlaThrAlaProAlaValLeuGlnSerLeuSerLeuLeuP 184
|||||
509 GTTCGGCTGGGCCCACTGCACCTGCTGCTGCAATCCCTCAGCCCTCTCT 558
```

```
184 heLeuProAlaGlyThrAspGluThrAlaThrHisLysAspLeuIlePro 200
|||||
559 TCCTCCCTGCTGCTACAGATGAGACTGCAACACACAAAGGACCTCATCCCA 608
|||||
201 LeuGlnGlyGlyGluAlaProLysLeuGlyProGlyArgProArgThrSe 217
|||||
609 CTCAGGGAGGTGAGGCCCCCAAGCTGGGCGCGGGAGGCCACGGTACTC 658
|||||
217 rPheLeuAspLeuPheArgAlaArgAspAsnMetArgGlyArgThrV 234
|||||
659 CTTTCTGGACCTCTTCAGGCGCAGCGATAAACATGCGAGCGCGGACACAG 708
|||||
234 aIGlyLeuGlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnVal 250
|||||
709 TGGGCGCTGGGGCTGGTCTCTTCCAGCAACTAACAGGGCAGGCCAACGCTG 758
|||||
251 LeuCysTyrAlaSerThrIlePheSerSerValGlyPheHisGlyCys 267
|||||
759 CTGTGCTATGCTCCACCATCTTCAGCTCCGTTGGTTTCCATGGGGGATC 808
|||||
267 rSerAlaValLeuAlaSerValGlyLeuGlyAlaValLysValAlaAlaT 284
|||||
809 CTCAGCGCTGCTGGCTCTCTGGGCTTGGCGCAGTGAAGTGGCAGCTA 858
|||||
284 hrLeuThrAlaMetGlyLeuValAspArgAlaGlyArgAlaLeuLeu 300
|||||
859 CCCTGACGCCCATGGGCTGGTGGACCGTGCAGCGCGCAGGGCTCTGTTG 908
|||||
301 LeuAlaGlyCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuVa 317
|||||
909 CTAGCTGGCTGTGCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 958
|||||
317 lSerPheAlaValProMetAspSerGlyProSerCysLeuAlaValProA 334
|||||
959 CAGCTTTGCGCTGCCCATGAGCTCAGGCCCAAGCTGCTGGCTGTGCCCA 1008
|||||
334 snAlaThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnAsp 350
|||||
1009 ATGCCACCGGGCAGACAGGCTCCCTGGAGACTCTGGCCTGCTGCGAGAC 1058
|||||
351 SerSerLeuProProIleProArgThrAsnGluAspGlnArgGluProI 367
|||||
1059 TCCTCTCTACCTCCCATTCACAGGACCAATGAGGACCAAGGAGGACCAAT 1108
|||||
367 eLeuSerThrAlaLysLysThrLysProHisProArgSerGlyAspProS 384
|||||
1109 CTTGTCCACTGCTAAGAAACCAAGCCCATCCACATCTGGAGACCCCT 1158
|||||
384 eAlaProProArgLeuAlaLeuSerSerAlaLeuProGlyProLeu 400
|||||
1159 CAGCCCTCTCGGCTGGGCTGAGCTCTGCCCTCCCTGGGCCCCCTCTG 1208
|||||
401 ProAlaArgGlyHisAlaLeuLeuArgTTrpThrAlaLeuLeuCysLeuMe 417
|||||
1209 CCGCTCGGGGGGATGACATGCTGCTCGGCTCGGACCGACCTGCTGCTGAT 1258
|||||
417 tValPheValSerAlaPheSerPheGlyPheGlyProValThrTrpLeuV 434
|||||
1259 GGTCTTTGTGTCAGTGCCTCTCTCTTTGGGTTTGGGCCAGTGACCTGGCTG 1308
|||||
434 aLeuSerGluIleTyrProValGluIleArgGlyArgAlaPheAlaPhe 450
|||||
1309 TCCTCAGCAGATCTACCTGTGGATACGAGGAAGAGCCCTTCGCTTC 1358
|||||
451 CysAsnSerPheAsnTrpAlaAlaAsnLeuPheIleSerLeuSerPheLe 467
|||||
1359 TGAACAGCTTCAACTGGGCGGCCAACCTCTTCATCAGGCTCTCTCTTCCT 1408
|||||
467 uAspLeuIleGlyThrIleGlyLeuSerTrpThrPheLeuLeuTyrGlyL 484
|||||
1409 CGATCTCATTCGACCATCGGCTTGTCTGGACCTTCTCTCTCTCTCTCTCT 1458
|||||
```

```
484 euThrAlaValLeuLeuGlyPheIleTyrLeuPheValProGluThr 500
|||||
1459 TGACCCCTCTCTCGGCTGGGCTTCATCTATTATTTCCTCTGAAACA 1508
|||||
501 LysGlyGlnSerLeuAlaGluIleAspGlnGlnPheGlnLysArgArgPh 517
|||||
1509 AAGGCCAGTCTGTTGCAGAGATAGACACAGCTTCCAGAAGACAGCGTT 1558
|||||
517 eThrLeuSerPheGlyHisArgGlnAsnSerThrGlyIleProTyrSerA 534
|||||
1559 CACCCTGAGCTTTGGCCACAGGCGAAGCTCCACTGGCATCCCGTACAGCC 1608
|||||
```

534 rgIleGluIleSerAlaAlaSer 541

1609 GCATCGAGATCTCTGGCGCTCC 1631

seq_name: gb_pr:AF321240

seq_documentation_block:

LOCUS AF321240 4167 bp mRNA PRI 25-MAR-2001
DEFINITION Homo sapiens facilitative glucose transporter GLUT10 (SLC2A10)
mRNA, complete cds.

ACCESSION AF321240

VERSION AF321240.1 GI:13445574

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 4167)

McVie-Wyllie, A.J., Lamson, D.R. and Chen, Y.T.

Molecular Cloning of a Novel Member of the GLUT Family of

Transporters, SLC2A10 (GLUT10), Localized on Chromosome 20q13.1: A

Candidate Gene for NIDDM Susceptibility

Genomics 72 (1), 113-117 (2001)

11247674

2 (bases 1 to 4167)

McVie-Wyllie, A.J., Lamson, D.R. and Chen, Y.T.

Direct Submission

Submitted (14-NOV-2000) Medical Genetics, Duke University Medical

Center, Trent Drive, Durham, NC 27710, USA

Location/Qualifiers

1. 4167

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/map="20q13.1"

1. 4167

/gene="SLC2A10"

53. 1678

/gene="SLC2A10"

/note="member of the facilitative glucose transporter

family"

/codon_start=1

/product="facilitative glucose transporter GLUT10"

/protein_id="AAK26294.1"

/db_xref="gi:13445574"

/translation="MGHSPVPVPLCASVSLGLGTFGYELAVISGALLPLQLDFGLSC

LEQFLVGLSLGLLGLASLVGGFLIDCYGRKOAILGSLNVLGLAGSLTGLAGSLAWLV

LGRAVGVPAISLSMACCIVVSELVGPQRGVLSLVFAGITVGTILLSYALNYALACT

PWGRHMFGNATAPVQLQSLFLPAGTDEATATKDLIPLOGGEANPLGPRYSF

LDLFRANDNRHRTTVGLGLVLPQGLTGPQNVLCYASTIFSSVGFHGGSSAVLASVGL

GAVKAATLTAMGLVDRAGRALLAGCALMALSVSGILGVSPFVPMDSGSPCLAVPN

ATGOTGLPDSGLQDSLSLPPRTNEDREPLSTAKTKTPHPSRSDGSPCLAVPN

SALPGPLPARGHALLRWATLCLMVFVSASFSGFVPVTLVLSIYVPEIRGAFAP

CNSEFWANLPLSLSLDLICTIGLSWTFLLYGLTAVLGLGFIYLPVETGQSLAEI

DQFQKRFTLSFGHRNSGTGIPSRISAS"

916 a 1090 c 991 g 1170 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 2765.00

Length: 541

Ratio: 5.111 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-652-292-2 x AF321240 ..
Align seg 1/1 to: AF321240 from: 1 to: 4167

1 MetGlyHisSerProProValLeuProLeuCysAlaSerValSerLeuLeu 17
|||||
53 ATGGGCCACTCCCACTGCTGCTTTGTGTGCTGCTGCTGCTGCTGCT 102
|||||
17 uClyGlyLeuThrPheGlyTyrGluLeuAlaValIleSerGlyAlaLeuL 34
|||||
103 GGGTGGGCTGACCTTGGTTATGAACATGGCAGTGCATATACAGGTGCCCTGC 152
|||||
34 euProLeuGlnLeuAspPheGlyLeuSerCysLeuGluGlnGluPheLeu 50
|||||
153 TGCCACTGCAGCTTGACTTTGGGCTAGCTGCTTTGGAGCAGGATTCCTG 202
|||||
51 ValGlySerLeuLeuGlyAlaLeuLeuAlaSerLeuValGlyGlyPh 67
|||||
203 GTGGCAGCCTGCTCCTGGGGCTCTCCTGCCCTCCCTGGTGGTGGCTT 252
|||||
67 eLeuIleAspCysTyrGlyArgLysGlnAlaIleLeuGlySerAsnLeuV 84
|||||
253 CCTCATTTGACTGCTATGGCAGGAAGCAAGCCATCCTCGGAGCAACTTGG 302
|||||
84 alLeuLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAlaTri 100
|||||
303 TGCTGCTGGCAGGAGCAGCTGACCTGGGCTGGCTGGTTCCTGGCCCTGG 352
|||||
101 LeuValLeuGlyArgAlaValValGlyPheAlaIleSerLeuSerMe 117
|||||
353 CTGGTCTGGGCGCGCTGTGGTGGCTTGGCCATTTCCCTCTCCTCCAT 402
|||||
117 tAlaCysCysTyrLeuValSerGluLeuValGlyProArgGlnArgGlyV 134
|||||
403 GGCTTGCTGATATCTACGTGTGACAGCTGGTGGGGCCACGCGCGGGAG 452
|||||
134 alLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeuSer 150
|||||
453 TGCTGGTGTCCCTCTATGAGCAGGCATCACCCTGGGCTGCTGCTCTCC 502
|||||
151 TyrAlaLeuAsnTyrAlaLeuAlaGlyThrProTrpGlyTrpArgHisMe 167
|||||
503 TATGCCCTCAACTATGCACTGGCTGTGTGTACCCCTGGGGATGGAGGCACAT 552
|||||
167 tPheGlyTrpAlaThrAlaProAlaValLeuGlnSerLeuSerLeuLeuP 184
|||||
553 GTTCGGCTGGCCACTGCACTGCTGCTGCTGCAATCCCTCAGCCTCTCT 602
|||||
184 heLeuProAlaGlyThrAspGluThrAlaThrHisLysAspLeuIlePro 200
|||||
603 TCCTCCCTGTGTGTACAGATGAGACTGCAACACACAAGGACCTCATCCCA 652
|||||
201 LeuGlnGlyGlyGluAlaProLysLeuGlyProGlyArgProArgTyrSe 217
|||||
653 CTCAGGGAGGTGAGGCCCCCAAGCTGGGGCCGGGAGGCGGACTC 702
|||||
217 rPheLeuAspLeuPheArgAlaArgAspAsnMetArgGlyArgThrThrV 234
|||||
703 CTTTCTGGACCTCTTCAGGGCAGCCGATACATGCGAGGCCGACACAG 752
|||||
234 alGlyLeuGlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnVal 250
|||||
753 TGGGCTGGGGCTGGTGGCTCTTCCAGCAACTAACAGGCGAGCCCAACGTG 802
|||||
251 LeuCysTyrAlaSerThrIlePheSerSerValGlyPheHisGlyGlySe 267
|||||
803 CTGTGTATGCTCCACCATCTTCAGCTCGTGGTTTCCATGGGGGATC 852
|||||
267 rSerAlaValLeuLeuSerValGlyLeuGlyAlaValLysValAlaAlat 284

|||||
853 CTCAGCCGTGCTGGCCTCTGTGGGGCTTGGCGCAGTGAAGTGGCAGCTA 902
|||||
284 hrLeuThrAlaMetGlyLeuValAspArgAlaGlyArgArgAlaLeuLeu 300
|||||
903 CCCTGACCGCCATGGGGCTGGTGGACCGTGCAGGCCGACAGGGCTCTGTTG 952
|||||
301 LeuAlaGlyCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuVa 317
|||||
953 CTAGCTGGCTGTGCCCTCATGGCCCTGTCCGTCACTGGCATAGGCCCTCGT 1002
|||||
317 lSerPheAlaValProMetAspSerGlyProSerCysLeuAlaValProA 334
|||||
1003 CAGCTTTGGCGTGGCCATGGACTCAGGCCCAAGCTGTCTGGCTGGCCCA 1052
|||||
334 snAlaThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnAsp 350
|||||
1053 ATGCCACCGGGCAGACAGGGCTCCCTGGAGACTCTGGCCTGCTGCAGGAC 1102
|||||
351 SerSerLeuProProIleProArgThrAsnGluAspGlnArgGluProIl 367
|||||
1103 TCCTCTCTACCTCCCATTTCCAAGGACCAATGAGGACCAAGGGAGCCCAAT 1152
|||||
367 eLeuSerThrAlaLysLysThrLysProHisProArgSerGlyAspProS 384
|||||
1153 CTTGTCCACTGCTAAGAAACCAAGCCCATCCAGATCTGGAGACCCCT 1202
|||||
384 erAlaProProArgLeuAlaLeuSerSerAlaLeuProGlyProLeu 400
|||||
1203 CAGCCCTCTCGGCTGGCCCTGAGCTCTGCCCTGCCCTGGGCCCCCTCTG 1252
|||||
401 ProAlaArgGlyHisAlaLeuLeuArgTyrThrAlaLeuLeuCysLeuMe 417
|||||
1253 CCCGCTGGGGGATGCACTGCTGGCTGGACCGCACTGCTGTGCTGAT 1302
|||||
417 tValPheValSerAlaPheSerPheGlyPheGlyProValThrTrpLeuV 434
|||||
1303 GGTCTTTGTGAGTGCCTTCTCTTTGGTGGGCCAGTGCCTGGGCTTG 1352
|||||
434 alLeuSerGluIleTyrProValGluLeuArgGlyArgAlaPheAlaPhe 450
|||||
1353 TCCTCAGCGAGATCTACCCCTGGAGATACGAGGAAGAGCCTTCGCCCTTC 1402
|||||
451 CysAsnSerPheAsnTrpAlaAlaAsnLeuPheIleSerLeuSerPheLe 467
|||||
1403 TGCAACAGCTTCAACTGGGGCGCAACCTCTTCATCAGCCTCTCCTTCT 1452
|||||
467 uAspLeuIleGlyThrIleGlyLeuSerTrpThrPheLeuLeuTyrGlyL 484
|||||
1453 CGATCTCATTTGGCACCATTGGCTGTGCTGGACCTTCTCTCTACGGAC 1502
|||||
484 euThrAlaValLeuGlyLeuGlyPheIleTyrLeuPheValProGluThr 500
|||||
1503 TGACCGCTGCTCTCGGCTGGGCTTCTATCTATTTATTTGTTCTGAAACA 1552
|||||
501 LysGlyGlnSerLeuAlaGluIleAspGlnGlnPheGlnLysArgArgPh 517
|||||
1553 AAAGGCCAGTCTGTCGAGAGATAGACCAGCAGTTCACAGAAAGAGACGTT 1602
|||||
517 eThrLeuSerPheGlyHisArgGlnAsnSerThrGlyIleProTyrSera 534
|||||
1603 CACCCTGAGCTTTGGCCACGAGGAGAACTCCACTGGCATCCCTGACAGCC 1652
|||||
534 rgIleGluIleSerAlaAlaSer 541
|||||
1653 GCATCGAGATCTCTGGGCTCC 1675

seq_name: gb_pr:AF248053

seq_documentation_block:
LOCUS AF248053 4396 bp mRNA
DEFINITION Homo sapiens glucose transporter (GLUT10) mRNA, complete cds.
ACCESSION AF248053

12-APR-2001
PRI

```

VERSION      AF248053.1  GI:13603726
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 4396)
AUTHORS      Fossey,S.C., Mihic,S.J., Craddock,A.L., Mychaleckyj,J.C.,
             Dawson,P.A. and Bowden,D.W.
TITLE        GLUT10: A novel glucose transporter in the type 2 diabetes linked
             region of chromosome 20q12-13.1
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 4396)
AUTHORS      Fossey,S.C., Mihic,S.J., Craddock,A.L., Mychaleckyj,J.C.,
             Dawson,P.A. and Bowden,D.W.
TITLE        Direct Submission
JOURNAL      Submitted (22-MAR-2000) Molecular Genetics, Wake Forest University,
             Medical Center Boulevard, Winston-Salem, NC 27106, USA
FEATURES     Location/Qualifiers
             1..4396
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="20"
             /map="20q12-q13.1"
             1..4396
             /gene="GLUT10"
             251..1876
             /gene="GLUT10"
             /codon_start=1
             /product="glucose transporter"
             /protein_id="AAK31911.1"
             /db_xref="GI:13603727"
             /translation="MGHSPVPVPLCASVSLGLGTFGYELAVISGALLPLQLDFGLSC
             LEQFLVGLSLGLALSLVGGFLIDCYRKQALIGSLNVLGLAGSLTLGLAGSLAWLV
             LGRVAVGFAISLSSMACCIYVSELVGPQRGVLSLYEAGITVGLISYALNYALAGT
             PWGRHFGWATAPAVLQSLSLFLPAGCTATATHKDLIPLQGGEPKLGPRPRYSP
             LDLEFRADNRKRTYVGLGLVFOOLQOPNVLCTYASTIFSVGFHGSSAVLASVGL
             GAVKVAATLTMGLVDRAGRALLLAGCALMALSVSGIGLSVAVPMDSGPSCLAVPN
             ATGQTGLDGLSLQDLSLPIPTNEDQREPIILSTAKTKPHRPSGDPSPAPRLALS
             SCLNCPPLPARGHALLRWLTALCLMVFSVAFSGFQPVTLVLSIYVPEIRGRAFAF
             CNSPNAANLPIISLSLDLICTIGLSWTFLLYGLTAVLGLGLFVLPVETKQSLAEI
             DQDFQKRFTLSFGHRNSTGIPYSRIEISAS"
BASE COUNT   939 a 1139 c 1102 g 1216 t
ORIGIN

alignment_scores:
  Quality: 2765.00      Length: 541
  Ratio: 5.111          Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-652-292-2 x AF248053

Align seg 1/1 to: AF248053 from: 1 to: 4396

1 MetGlyHisSerProProValLeuProLeuCysAlaSerValSerLeuLeu 17
|||||
251 ATGGGCCACTCCACCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300

17 uGlyGlyLeuThrPheGlyTyrGluLeuAlaValIleSerGlyAlaLeuL 34
|||||
301 GGTGGCCCTGACCTTTGGTTATGAACGTGGCAGTCATATCAGTGGCCCTG 350

34 euProLeuGlnLeuAspPheGlyLeuSerCysLeuGluGlnPheLeu 50
|||||
351 TGGCACTGACGCTGACTTTGGGCTGAAGCTGCTGGAGCAGGAGTTCCTG 400

51 ValGlySerLeuLeuLeuGlyAlaLeuLeuAlaSerLeuValGlyGlyP 67
|||||
401 GTGGGCGAGCTGCTCTGGGGGCTCTCTCTCGCTCCCTGCTGCTGGTGG 450

67 eLeuIleAspCysTyrGlyArgLysGlnAlaIleLeuGlySerAsnLeu 84
|||||

```

```

|||||
451 CCTCATTTGACTGCTATGGCAGGAAGCAAGCATCTCTCGGAGCAACTTGG 500

84 alLeuLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAlaTrp 100
|||||
501 TGTGCTGGCAGCAGCAGCTGACCTGGCGCTGGCTGCTGCTGCTGCTGCTG 550

101 LeuValLeuGlyArgAlaValValGlyPheAlaIleSerLeuSerSerMe 117
|||||
551 CTGCTCTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600

117 tAlaCysCysIleTyrValSerGluLeuValGlyProArgGlnArgGlyV 134
|||||
601 GGCTTGTGCTGCTATCTACGCTGTCAGAGCTGGTGGGCGCACGCGGGGAG 650

134 alLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeuSer 150
|||||
651 TGTGCTGCTGCTCTATGAGCAGCAGCATCACCTGGGCGCATCTGCTCTCC 700

151 TyrAlaLeuAsnTyrAlaLeuAlaGlyThrProTrpGlyTrpArgHisMe 167
|||||
701 TATGCCCTCAACTATGCATGCTGGTGGTACCCCTGGGGATGGAGGCACAT 750

167 tPheGlyTrpAlaThrAlaProAlaValLeuGlnSerLeuSerLeuLeuP 184
|||||
751 GTTGGGCTGGGCGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 800

184 heLeuProAlaGlyThrAspGluThrAlaThrHisLysAspLeuIlePro 200
|||||
801 TCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850

201 LeuGlnGlyGlyGluAlaProLysLeuGlyProGlyArgProArgTyrSe 217
|||||
851 CTCAGGAGGTGAGGCGCCCAAGCTGGCGCGGGAGGCGGACGCTACTC 900

217 rPheLeuAspLeuPheArgAlaArgAspAsnMetArgGlyArgThrThrV 234
|||||
901 CTTTCTGGAGCTCTTTCAGGCGACGCGATAACATCGAGGCGGACACAG 950

234 alGlyLeuGlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnVal 250
|||||
951 TGGGCTGGGCGCTGGTCTCTTCCAGCACTAACAGGCGGACCCCACTG 1000

251 LeuCysTyrAlaSerThrIlePheSerSerValGlyPheHisGlyGlySe 267
|||||
1001 CTGTGCTATGCTCCACCATCTTTCAGCTGCTGCTGCTGCTGCTGCTG 1050

267 rSerAlaValLeuAlaSerValGlyLeuGlyAlaValIleValAlaAla 284
|||||
1051 CTCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1100

284 hrLeuThrAlaMetGlyLeuValAspArgAlaGlyArgArgAlaLeuLeu 300
|||||
1101 CCCTGACCGCATGGGCGCTGGTGGACCGCTGAGCGCGCAGGGCTCTGT 1150

301 LeuAlaGlyCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuVa 317
|||||
1151 CTAGCTGGCTGGCGCTCATGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

317 lSerPheAlaValProMetAspSerGlyProSerCysLeuAlaValProA 334
|||||
1201 CAGCTTTGCGCTGCGCATGGACTGAGCCCAAGCTGCTGGCTGTGCCCA 1250

334 snAlaThrGlyGlnThrGlyLeuProGlyLeuAspSerGlyLeuGlnAsp 350
|||||
1251 ATGCCACCGGCGACAGAGGCTCTCTGAGACTCTGGCCTGCTGCTGCTG 1300

351 SerSerLeuProProIleProArgThrAsnGluAspGlnArgGluProIl 367
|||||
1301 TCCTCTCTACCTCCCATTTCCAGGACCAATCAGGACCAAGAGGAGCCA 1350

367 eLeuSerThrAlaLysLysThrLysProHisProArgSerGlyAspProS 384
|||||

```

1351 CTTGTCCACTCCTAGAAAACCAAGCCCCCATCCAGATCTCGAGACCCCT 1400
 384 erAlaProArgLeuAlaLeuSerSerAlaLeuProGlyProProLeu 400
 1401 CAGCCCTCCTCGCTGCGCTGAGCTGCTCCCTCCCTGCGCCCTCTG 1450
 401 ProAlaArgGlyHisAlaLeuLeuArgTrpThrAlaLeuLeuCysLeu 417
 1451 CCCGCTCGGGGCGCATGCACTGCTCGCTGGACCGCACTGCTGCTGAT 1500
 417 tValPheValSerAlaPheSerPheGlyPheGlyProValThrTrpLeu 434
 1501 GGTCTTGTAGTGCCTCTCTCTTGGGTTGGGCGAGTGACCTGGCTG 1550
 434 alLeuSerGluIleTyProValGluIleArgGlyArgAlaPheAlaPhe 450
 1551 TCCTCAGGAGATCTACCTCTGGAGATACGAGGAGAGCGCTTCGCTTC 1600
 451 CysAsnSerPheAsnTrpAlaAlaAsnLeuPheLeSerLeuSerPheLe 467
 1601 TGCACAGCTTCAACTGGGCGGCAAGCTCTTCATCAGCCCTCCTCCT 1650
 467 uAspLeuIleGlyThrIleGlyLeuSerTrpThrPheLeuLeuTyGlyL 484
 1651 CGATCTCATTTGGCACCATCGCTTGTCTCGACCTTCCTGCTACGGAC 1700
 484 euThrAlaValLeuGlyLeuGlyPheIleTyLeuPheValProGluThr 500
 1701 TGACCGCTGCTCTGGCGCTGGCTTCATCTATTTATTTGTTCTTGAACA 1750
 501 LysGlyClnSerLeuAlaGluIleAspGlnGlnPheGlnLysArgArgPh 517
 1751 AAAGCCAGCTGCTGGCAGAGATAGACGAGCAGTTCAGAGAGACGGTT 1800
 517 eThrLeuSerPheGlyHisArgGlnAsnSerThrGlyIleProTySerA 534
 1801 CACCTCAGCTTGGCCACAGGAGAACTCCACTGGCATCCCGTACAGCC 1850
 534 rgIleGluLeuSerAlaAlaSer 541
 1851 GCATCAGATCTCTGCGGCTCC 1873

seq_name: gb_pr:HS28H20

seq_documentation_block: 19-APR-2001 PRI
 LOCUS HS28H20 127418 bp DNA
 DEFINITION Human DNA sequence from clone RPL-28H20 on chromosome 20q13.1
 Contains the SLC2A10 gene encoding a solute carrier family 2
 (facilitated glucose transporter) member 10, the 5' end of a novel
 gene, ESTs, STSS, GSSs and three CpG islands, complete sequence.
 AL031055
 VERSION AL031055.1 GI:4375937
 KEYWORDS HTG; CpG island; SLC2A10.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 127418)

REFERENCE
 AUTHORS Ramsay, H.
 JOURNAL Direct Submission
 CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerquests@sanger.ac.uk

COMMENT
 Submitted (18-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 On Mar 7, 1999 this sequence version replaced gi:4056528.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 20, constructed by the Sanger Centre Chromosome 20
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr20
 This sequence is the entire insert of clone RPL-28H20 This sequence
 was finished as follows unless otherwise noted: all regions were
 either double-stranded or sequenced with an alternate chemistry or
 covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by at least one
 plasmid subclone or more than one M13 subclone; and the assembly
 was confirmed by restriction digest. RPL-28H20 is from the library
 RPL-1 constructed by the group of Pieter de Jong. For further
 details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pCYPAC2.

FEATURES

source
 1..127418
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="q13.1"
 /clone="RPL-28H20"
 /clone_lib="RPL-1"
 complement(join{..1..562,2463..3110})
 /gene="dJ101A2.2"
 /note="match: cDNAs: Em:AK013049 Em:AK011574
 match: ESTs: Em:AA233726 Em:T87841 Em:AA756018 Em:AA113277
 Em:AA985248 Em:AA682750 Em:AA461487 Em:AI028588
 Em:AA659710 Em:AI149891 Em:AA579641 Em:AA516077
 Em:AA460132 Em:AA483490 Em:AA046575"
 /evidence="not_experimental"
 /product="dJ28H20.2 (novel protein)"
 complement(1..3110)
 /gene="dJ101A2.2"
 18..494
 /note="match: STS: Em:HS28H20T"
 complement(join{84..562,2463..2745})
 /gene="dJ101A2.2"
 /note="supported by GENESH and GENSCAN
 continues in Em:AL133520 as dJ101A2.2"
 /codon_start=1
 /evidence="not_experimental"
 /product="dJ28H20.2 (novel protein)"
 /protein_id="CAC00561.1"
 /db_xref="GI:9388402"
 /db_xref="SPTREMBL:O9NOE6"
 /translation="MAAARATTPADGEEPAEALAAARSSRPLSLGLVKGQAE
 ARVFRGFGRAAVIKHRPKGYRHPALERLGRRTVQEARLLRCRAGISAPVVE
 FVDYASNCLYMEIEIGSVTRDYIOSTMETKTPQGLSLAKTIQGLARMDEDLIH
 GDLTTSNMLLKPPLEQLNVLIDFGLSFISALPDKGVLYLEKAFLETHPNTETVF
 EAFKASYSTSSKKARPVLKLDVLEGRKRSWVG"
 867..1145
 /note="AluX repeat: matches 1..298 of consensus"
 1251..1353
 /note="MIR repeat: matches 50..151 of consensus"
 1672..1797
 /note="MIR repeat: matches 8..135 of consensus"
 1855..2079
 /note="MIR repeat: matches 33..262 of consensus"
 2155..2248
 /note="MIR repeat: matches 33..135 of consensus"
 2384..3150
 /note="CpG island"
 /evidence="not_experimental"
 2812..2871
 /note="3 copies 20 mer 85% conserved"
 complement(2876..3110)
 /gene="dJ101A2.2"
 /note="match: STS: Em:G15621"
 3865..3962

repeat_region

```

repeat_region /note="MER21B repeat: matches 299. .391 of consensus"
3963. 5655
repeat_region /note="L1M8A repeat: matches 4532. .6270 of consensus"
5669. 5895
repeat_region /note="L1ME repeat: matches 5501. .5739 of consensus"
6332. 6548
repeat_region /note="L2 repeat: matches 2524. .2746 of consensus"
6626. 6915
repeat_region /note="Alusx repeat: matches 3. .290 of consensus"
6918. 6980
repeat_region /note="MER94 repeat: matches 1. .62 of consensus"
7023. 7322
repeat_region /note="AluJo repeat: matches 1. .312 of consensus"
7323. 7417
repeat_region /note="L2 repeat: matches 2389. .2513 of consensus"
7457. 7623
repeat_region /note="MIR repeat: matches 93. .262 of consensus"
7864. 7934
repeat_region /note="MLT1I repeat: matches 311. .382 of consensus"
8046. 8240
repeat_region /note="MLT1J repeat: matches 9. .211 of consensus"
8347. 8647
repeat_region /note="Alusq repeat: matches 1. .300 of consensus"
8685. 8793
repeat_region /note="MLT1J repeat: matches 271. .366 of consensus"
9095. 9221
repeat_region /note="MIR repeat: matches 102. .242 of consensus"
9647. 9776
repeat_region /note="L2 repeat: matches 2170. .2303 of consensus"
9793. 10028
repeat_region /note="AluJo repeat: matches 74. .306 of consensus"
10038. 10206
repeat_region /note="LTR16C repeat: matches 126. .316 of consensus"
10631. 10690
repeat_region /note="MLT1D repeat: matches 445. .505 of consensus"
10691. 10995
repeat_region /note="AluJb repeat: matches 1. .295 of consensus"
10996. 11200
repeat_region /note="MLT1D repeat: matches 248. .445 of consensus"
11232. 11634
repeat_region /note="MSTB repeat: matches 2. .409 of consensus"
11670. 11761
repeat_region /note="MLT1D repeat: matches 128. .192 of consensus"
11762. 12062
repeat_region /note="Alusx repeat: matches 1. .293 of consensus"
12063. 12180
repeat_region /note="MLT1D repeat: matches 1. .128 of consensus"
complement(12181. 12344)
misc_feature /note="match: GSS: Em:AQ190762"
13011. 13129
repeat_region /note="MLT1J repeat: matches 252. .368 of consensus"
13210. 13392
repeat_region /note="MER20 repeat: matches 1. .215 of consensus"
13397. 13442
repeat_region /note="MIR repeat: matches 104. .146 of consensus"
13660. 13783
repeat_region /note="LTR33 repeat: matches 330. .452 of consensus"
13894. 14254
repeat_region /note="L2 repeat: matches 2255. .2619 of consensus"
14305. 14435
repeat_region /note="L2 repeat: matches 2617. .2748 of consensus"
14446. 14525
repeat_region /note="4 copies 20 mer 76% conserved"
15476. 15752
repeat_region /note="AluJb repeat: matches 2. .278 of consensus"
15774. 15952
repeat_region /note="MER58A repeat: matches 42. .220 of consensus"
16190. 16304
repeat_region /note="L2 repeat: matches 2657. .2744 of consensus"
16305. 16599
repeat_region /note="AluJb repeat: matches 1. .296 of consensus"
16600. 16635
repeat_region /note="L2 repeat: matches 2617. .2657 of consensus"

```

```

repeat_region 16641. 16803
/note="L1MA10 repeat: matches 6162. .6284 of consensus"
repeat_region 16804. 17190
/note="MER57B repeat: matches 1. .403 of consensus"
repeat_region 17198. 17297
/note="5 copies 20 mer 71% conserved"
repeat_region 17204. 17297
/note="47 copies 2 mer ta 70% conserved"
repeat_region 17559. 17868
/note="Alusx repeat: matches 1. .312 of consensus"
repeat_region 17939. 18240
/note="LTR26 repeat: matches 246. .553 of consensus"
misc_feature complement(18333. 18408)
/note="match: STS: Em:G42885"
repeat_region 18451. 18615
/note="FAM repeat: matches -126. .41 of consensus"
repeat_region 19258. 19336
/note="MER57-internal repeat: matches 7168. .7244 of consensus"
repeat_region 19343. 19485
/note="MER53 repeat: matches 2. .180 of consensus"
repeat_region 19721. 20190
/note="LTR8 repeat: matches 251. .691 of consensus"
repeat_region 20238. 20493
/note="LTR8 repeat: matches 1. .256 of consensus"
repeat_region 20704. 20978

alignment_scores:
  Quality: 2286.50      Length: 651
  Ratio: 4.855          Gaps: 2
  Percent Similarity: 72.350  Percent Identity: 72.197

alignment_block:
US-09-6521292-2 x HS28H20 ..

Align seg 1/1 to: HS28H20 from: 1 to: 127418

1 MetGlyHisSerProProValLeuProLeuCysAlaSerValSerLeuLe 17
:::|||||
38368 TTAGGCCACCTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 38417

17 uGlyGlyLeuThrPheGlyTyrGluLeuAlaValIleSerGlyAlaLeu 34
|||||
38418 GGGTGGCTGACCTTTGGTATGAACGGCAGTCATATCAGTGCCTTC 38467

34 euProLeuGlnLeuAspPheGlyLeuSerCysLeuGluGlnPheLeu 50
|||||
38468 TGGCAGCTGCAGCTTGACTTTGGGCTTAAGCTGCTTGGAGCAGAGTTCCTG 38517

51 ValGlySerLeuLeuGlyAlaValAlaLeuAlaSerLeuValGlyGly 67
|||||
38518 GTGGGAGCCTGCTCCTGGGGCTCTCCTGCTCCTCCTGCTGCTGCTT 38567

67 eLeuIleAspCysTyrGlyArgLysGlnAlaIleLeuGlySerAsnLeu 84
|||||
38568 CCTCATTTGACTGCTATGGCAGGAAGCAAGCATCTCGGGAGCACTGG 38617

84 allLeuLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAla 100
|||||
38618 TGCTGCTGGCAGGAGCCTGACCTGGGCTGGGCTGGCTGGCTGGCTG 38667

101 LeuValLeuGlyArgAlaValAlaValGlyPheAlaIleSerLeuSer 117
|||||
38668 CTGGTCTGGGCGCGCTGCTGGTGGCTTGGCTTGGCTTCCCTCTCCT 38717

117 tAlaCysCysIleTyrValSerGluLeuValGlyProArgGlnArgGly 134
|||||
38718 GGCTTGCTGTATCTACGTGTGACAGCTGCTGGGCGCAGCGAGGGGAG 38767

134 allLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuSer 150
|||||
38768 TGGTGGTGTCCCTCTATGAGGAGGAGCATCACCGTGGGCATCCTGCT 38817

```

151 TyrAlaLeuAsnTyrAlaLeuAlaGlyThrProTrpGlyTrpArgHisMe 167
|||||
38918 TATGCCCTCAACTATGCACTGGTGTGTACCCCTGGGGATGGAGGCACAT 38867
167 tPheGlyTrpAlaThrAlaProAlaValLeuInSerLeuSerLeuLeuP 184
|||||
38868 GTTCGGCTGGGCACATGCATCTGTCTGCAATCCCTCAGCCCTCCTCT 38917
184 heLeuProAlaGlyThrAspGluThrAlaThrHisLysAspLeuIlePro 200
|||||
38918 TCCTCCCTGCTGTGTACAGTACGACTGCAACACACAGGACCTCATCCCA 38967
201 LeuGlnGlyGluAlaProLysLeuGlyProGlyArgProArgTyrSe 217
|||||
38968 CTCAGGGAGGTGAGGCCCCCAAGCTGGGCCCGGGAGGCACGGTACTC 39017
217 rPheLeuAspLeuPheArgAlaArgAspAsnMetArgGlyArgThrTrv 234
|||||
39018 CTTTCTGGACCTCTTACGGGCACGCGATAACATGCGAGGCGCGACACAG 39067
234 aGlyLeuGlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnVal 250
|||||
39068 TGGCCTGGGGCTGGTCTCTTCAGCACTAACAGGGCAGCCCAACGTG 39117
251 LeuCysTyrAlaSerThrIlePheSerSerValGlyPheHisGlyGlySe 267
|||||
39118 CTGTCTATGCCCTCCACCATCTTCAGCTCGGTGGTTTCCATGGGGGATC 39167
267 rSerAlaValLeuAlaSerValGlyLeuGlyAlaValLysValAlaAlaT 284
|||||
39168 CTCAGCGTCTGGCTCTGTGGGCTTGGCGAGTGAAGGTGGCAGCTA 39217
284 hrLeuThrAlaMetGlyLeuValAspArgAlaGlyArgArgAlaLeuLeu 300
|||||
39218 CCCTGACCGCATGGGGCTGTGGACCGTGCAGGCCGCGAGGCTCTGTG 39267
301 LeuAlaGlyCysAlaLeuMetAlaLeuSerValSerIleGlyLeuVa 317
|||||
39268 CTAGCTGGCTGTGCCCTCATGGCCCTGTCCGTGAGTGGCATGGCCTCGT 39317
317 lSerPheAlaValProMetAspSerGlyProSerCysLeuAlaValProA 334
|||||
39318 CAGCTTTGGCTGCCATGACTCAGGCCCAAGCTGTCTGGCTGTGGCCCA 39367
334 snAlaThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnAsp 350
|||||
39368 ATGCCACGGGCAGACAGGCTCCCTGGAGACTCTGGCCTGTGCAGGAC 39417
351 SerSerLeuProPheProAlaThrAsnGluAspGlnArgGluProII 367
|||||
39418 TCCTCTCTACCTCCCATTTCCAGGACCAATGAGGACCAAGGGAGCCAAT 39467
367 eLeuSerThrAlaLysLysThrLysProHisProArgSerGlyAspProS 384
|||||
39468 CTTGTCCACTGCTTAGAAAACCAAGCCCCATCCAGATCTGGAGACCCCT 39517
384 erAlaProProArgLeuAlaLeuSerSerAlaLeuProGlyProProLeu 400
|||||
39518 CAGCCCTCTCCCTGGCTGGCCCTCAGCTCTGCCCTCCCTGGGCCCTCTG 39567
401 ProAlaArgGlyHisAlaLeuLeuArgTrpThrAlaLeuLeuCysLeuMe 417
|||||
39568 CCGCCTCGGGGGATGCACCTGCTGCGCTGGAGCCGACATGCTGTGCCTGAT 39617
417 tValPheValSerAlaPheSerPheGlyPheGly..... 428
39618 GGCTTTTGTGAGTGCCTTCTCCTTTGGGTTGGCCAGGTAAGTGGAGTT 39667
428 428
39668 TTCTTGAGGTGACTCTGGAGACTTCTACCCCTCCTAGGGGGAAGTTGG 39717
428 428

39718 GGACTTCCCACTGATCACCTGCGAGGTGTCTAGTGGGCGACTAGAGGT 39767
428 428
39768 GGGTTGACCATGAAGCCCTCAGGGCCCTCCTACCTACCTGCGATTGCTCCTC 39817
428 428
39818 CCAGGGCCAGTGCCTAATTTTATGTTGTAATTTGTTATTTCTTAAAGCC 39867
428 428
39868 TCCCAAGTCAATTTGAGCTTCAGGTCCCTTAAAGCCCGGATAGCTCAAA 39917
428 428
39918 CTGGAAGGAATTTAATGAATAGCTGAAACCAATCATCTCCAGTTCGTCC 39967
428 428
39968 AAGACTGTTCCTCCATCTCGGCATTGAAAGTCTCATGTCTCAGAAAACCTCT 40017
428 428
40018 CAGTCTTTCAGCAAACTGAGACTCTACTGTAAATTTCTTATAGCAGATGGGA 40067
428 428
40068 GTGGGAGTTTAGTATCAGGAGCCTGACACATAGAAATATCCCATTTGTTGAGA 40117
428 428
40118 GGCTGCATGTTTGACCTGATGGTCCAGGCCCGCCAGTCCACACAGCCC 40167
429 Pro..ValThrTrpLeuValLeuSerGlu 437
40168 AGGAGCCCTCTCTGCTCTCCACCCCTAGTAGCTGTGCTGTCTCAGCGAG 40217
438 lIleTyrProValGluIleArgGlyArgAlaPheAlaPheCysAsnSerPh 454
|||||
40218 ATCTACCTCTGGAGATACGAGGAAGAGCCTTCGCCCTTCTGCAACAGCTT 40267
454 eAsnTrpAlaAlaAsnLeuPheIleSerLeuSerPheLeuAspLeuIleG 471
|||||
40268 CAACTGGGCGGCCAACTCTTTCATCAGCCTCTCTCTCTCGATCTCATTTG 40317
471 ly 471
40318 GT 40319
seq_name: gb_htg:AL591064
seq_documentation_block:
LOCUS AL591064 200125 bp DNA HTG 03-MAY-2001
DEFINITION Mus musculus chromosome 2 clone RP23-395E18, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL591064
VERSION AL591064.3 GI:13990651
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 200125)
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (02-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On May 7, 2001 this sequence version replaced gi:13990284.
----- Genome Center
Center: Sanger Centre

Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquyesanger.ac.uk
 ----- Project Information
 Center project name: BM395E18
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 195932 bases at least Q40
 Consensus quality: 196830 bases at least Q30
 Consensus quality: 197490 bases at least Q20
 Insert size: 198525; sum-of-contigs
 Insert size: 208466; 5.9% error; agarose-fp
 Quality coverage: 8.60x in Q20 bases; sum-of-contigs Quality
 coverage: 8.32x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

source	Location/Qualifiers
1. 200125	Location/Qualifiers
1. 200125	1. 200125
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/chromosome="2"
	/clone="RP23-395E18"
	/clone_lib="RPCI-23"
1. 4470	1. 4470
	/note="assembly_fragment:01119"
	fragment_chain:1
	clone_end:T7
	vector_side:left
4571..16275	4571..16275
	/note="assembly_fragment:03729"
	fragment_chain:1
16376..33874	16376..33874
	/note="assembly_fragment:02569"
	fragment_chain:1
33975..46195	33975..46195
	/note="assembly_fragment:05190"
	fragment_chain:1
46296..57940	46296..57940
	/note="assembly_fragment:03914"
	fragment_chain:1
58041..62650	58041..62650
	/note="assembly_fragment:03121"
	fragment_chain:1
62751..71940	62751..71940
	/note="assembly_fragment:01509"
	fragment_chain:1
72041..80555	72041..80555
	/note="assembly_fragment:02606"
	fragment_chain:1
80656..85958	80656..85958
	/note="assembly_fragment:01353"
	fragment_chain:1
86059..113228	86059..113228
	/note="assembly_fragment:04226"
	fragment_chain:1
11329..122676	11329..122676
	/note="assembly_fragment:00791"
	fragment_chain:1
122777..153882	122777..153882
	/note="assembly_fragment:03653"
	fragment_chain:1
153983..162005	153983..162005
	/note="assembly_fragment:02332"
	fragment_chain:1
162106..184926	162106..184926
	/note="assembly_fragment:04610"
	fragment_chain:1
185027..190517	185027..190517

	/note="assembly_fragment:01620"
	fragment_chain:1
190618..196019	190618..196019
	/note="assembly_fragment:00987"
	fragment_chain:2
196120..200125	196120..200125
	/note="assembly_fragment:00065"
	fragment_chain:2
	clone_end:SP6
	vector_side:right
BASE COUNT	50402 a 48944 c 49953 g 49224 t 1602 others
ORIGIN	
alignment_scores:	
Quality:	1678.00 Length: 649
Ratio:	4.014 Gaps: 4
Percent Similarity:	64.407 Percent Identity: 56.394
alignment_block:	
US-09-652-292-2 x AL591064	..
Align seg 1/1 to: AL591064 from: 1 to: 200125	
2 GlyHisSerProValLeuProLeuCysAlaSerValSerLeuLeuG1 18	
163943 GGCCTTCGCCCAAGCTGCTCCCTGCTCTGTGCTCTGTGCTTTGGTGGG 163992	
18 yGlyLeuThrPheGlyTyrGluLeuAlaValIleSerGlyAlaLeuLeup 35	
163993 GGCCCTGACCTTCGGATATGAGCTGGCCGTCATATCTGGTGGCCCTCTCTAC 164042	
35 roLeuGlnLeuAspPheGlyLeuSerCysLeuGluGlnGluPheLeuVal 51	
164043 CACTCCAGCTGAACCTTCGGGCTGAGTTGCCTCGAACAGGAGCTCTGGTG 164092	
52 GlySerLeuLeuLeuGlyAlaLeuAlaLeuAlaSerLeuValGlyGlyPheLe 68	
164093 GGCAGTCTGCTCTTTGGTGGTCTCTCCCTGCTCTCTCTGGGGGGCTTCCT 164142	
68 uileAspCysTyrGlyArgLysGlnAlaIleLeuGlySerAsnLeuValL 85	
164143 CATTGACTGCTATGGCAGGAGACGAGCCATCTTAGGAGCAATGCCGTGC 164192	
85 euLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAlaTrpLeu 101	
164193 TGCTGGCTGGCAGCCTGATTCTGGGTCTGGCCAGCTCCCTGGCCTGGCTG 164242	
102 valLeuGlyArgAlaValValGlyPheAlaIleSerLeuSerSerMetAl 118	
164243 CTCCTGGGCGGTTTATCTGCTGGCTTTGGCATCTCTGCTTCCATGGC 164292	
118 aCysCysIleTyrValSerGluLeuValGlyProArgGlnArgGlyValL 135	
164293 TTGCTGTATCTAGCTGTACAGATTGGTGGGCGCGGAGAGGGGTGTCG 164342	
135 euValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeuSerTyr 151	
164343 TGCTGTCCCTCTACGAGGTGGGCATCCTCTGGGCATTCTGTCTCTTAC 164392	
152 AlaLeuAsnTyrAlaLeuAlaGlyThrProTrpGlyTrpArgHisMetPh 168	
164393 GGCCTCAACTACGCTCTGGCTGGCAGCCCTTGGGGCTGGAGGCACATGTT 164442	
168 eGlyTrpAlaThrAlaProAlaValLeuGlnSerLeuSerLeuPheL 185	
164443 TGCTGGGAGCTGACCTGCTCTCTACAGTCACTCAGCCTCTTCTTTC 164492	
185 euProAlaGlyThrAspGluThrAlaThrHisLysAspLeuIleProLeu 201	
164493 TCCTCTGCTGGTGCAGAGGGCAGCAGCAGCCCAAGACCTTATCCCATC 164542	
202 GlnGlyGlyGluAlaProLysLeuGlyProGlyArgProArgTyrSerPh 218	

164543 CAGGGAAGGGAGACAAGCAAAACCGGGCCCTGGTGAAGCCACACATACAGCTT 164592
164592
164593 eLeuAspLeuPheArgAlaArgAspAsnMetArgGlyArgThrThrValG 235
164593 TCTGGACCTTTCAGGGCCAGCAGCATGTGGAGCCGAGCTGTAGTGG 164642
235 lLeuGlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnValLeu 251
164643 GGCTGGGCTGTGTCTTTCAGCAGCTCACAGGACAGCCCAATGTGTG 164692
252 CysTyrAlaSerThrIlePheSerSerValGlyPheHisGlyGlySerSe 268
164693 TATTACGGCTTACCATCTTTCGCTCCGCTGGCTTCCAGGGGGCTCCTC 164742
268 rAlaValLeuAlaSerValGlyLeuGlyAlaValValLysValAlaAlaThrL 285
164743 AGCTGTGTGCTTCTGTGGGCTTGGCACGCTGAAGCTGGCTGCCACT 164792
285 euThrAlaMetGlyLeuValAspArgAlaGlyArgAlaLeuLeuLeu 301
164793 TGGTGGCCAGGGGCTGTGTGACCGTGCAGCCGAGAGCTTCTCTGCTC 164842
302 AlaGlyCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuValSe 318
164843 TTTGGATGGCTCTCATGGCCCTTATCGGTACGGCATAGGCGCTGGTCAG 164892
318 rPheAlaValProMetAspSerGlyProSerCysLeuAlaValProAsnA 335
164893 CTTTGTGTGTCTTCTGGACTGTGACCCAGCTGCTGGCCACATCCAATG 164942
335 laThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnAspSer 351
164943 CCAGCCACAGGTGGATGTGCTCGGAAGCTCGGTTTGTGTGCTGTCA 164992
352 SerLeuProIleProArgThrAsnGluAspGlnArgGluProIleLe 368
164993 TCTCTGGCGGCAGTGTCTACACCAANTGGGACCAAGCAGCTGGTCTC 165042
368 uSerThrAlaLysLysThrLysProHisProArgSerGlyAspProSerA 385
165043 GTCACTGACTGAGAGA...CCCATCCATCCAGTCATCACAGCATCCCTGG 165089
385 laProProArgLeuAlaLeuSerSerAlaLeuProGlyProProLeuPro 401
165090 GACCA.....GTCTGAACACTGCCCTCCCGCATTCOCACCAAGTCCC 165130
402 AlaArgGlyHisAlaLeuLeuArgTrpThrAlaLeuLeuCysLeuMetVa 418
165131 ATCCTTGAGCATACCTGTGTGTGTCTCGCTGGTGTTCATGATGGT 165180
418 lPheValSerAlaPheSerPheGlyPheGlyPro.Val..... 430
165181 CTACGTGAGCGCTCTCTCGTTGGCTTGGACAGGTAGGAGGCTTATC 165230
430 430
165231 GGGTAGGGGAATCTAGAGCCTTCTGTCCCTTTCAGTGGGCTTCTTAGTC 165280
430 430
165281 TAGGGACACTCAGAGCTTACTGATGGCAGCGGAGGCTCAGGGTTCCTT 165330
430 430
165331 GCTTATATCAAGGGGCTCCCCATAAGTTTAAATCCAACACATCTGCAA 165380
430 430
165381 GTTTTAAAGCAACCTCCCTCTGCCGTTGTGGGCAACCACTTAAA 165430
430 430

165431	CTCCAGGTCTCTTCAAACCTTAGACCTATCCTCGTGGTTCATTGAA	165480
430	430
165481	GGCTGGCTTGAAGAGATGTAAACAGTCAGATAAAACCAGTCATACCA	165530
430	430
165531	GTTTGCCTGGGGGGGGGGCCCTTCTGCTTGGCATGCAATGTCTGTG	165580
430	430
165581	TCCAGGAACCTGAGACTGCCCTATTTCATGGAGTTCCTCCAGGGGCACA	165630
430	430
165631	GAAGTTGTGCACTCTGAGAGAACTCTGGGCCAAGAGCATGACACAGGTCTC	165680
430	430
165681	AGTTAAACGAGGTTGGGAGGTGGCTGTCCAGTTCATGGGTGCCCAAGATC	165730
431ThrTrpLeuValLeuSerGluIleTyrP	440
165731	TCTCTTGTCTCTCGCTCCAGTAACCTGGTGGTCTCTAGTCAGATCTACC	165780
440	rovalGluIleAArgGlyArgAlaPheAlaPheCysAsnSerPheAsnTrp	456
165781	CAGCGGAGATCCGAGGGAGACCTTCGGCTTCTGCAGCAGCTTCAACTGG	165830
457	AlaAlaAsnLeuPheIleSerLeuSerPheLeuAspLeuIleGly	471
165831	GCAGCTAACCTCTTTATCAGGCTCTCCTTCTCGACCTGATCGGT	165875

seq_name: qb_htq:AL591805

seq_documentation_block:

LOCUS	AL591805	243075 bp	DNA	HTG	22-JUL-2001
DEFINITION	Mus musculus chromosome 2 clone RP23-90N15, *** SEQUENCING IN PROGRESS ***, in unordered pieces.				

ACCESSION AL591805

VERSION AL591805.9 GI:14626203

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; COURSE

SOURCE house mouse.

ORGANISM *Mus musculus*
Eukaryota. M

REFERENCE
1 (bases 1 to 243075)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia

AUTHORS Sims, S.

TITLE	Direct Submission
1. <i>Abstract</i>	
2. <i>Introduction</i>	
3. <i>Methods</i>	
4. <i>Results</i>	
5. <i>Discussion</i>	
6. <i>Conclusion</i>	
7. <i>References</i>	
8. <i>Appendix</i>	
9. <i>Supplementary Materials</i>	
10. <i>Other</i>	

JOURNAL Submitted (21-JUL-2010)

CB10 ISA, UK. E-mail: cb10@isa.co.uk

COMMENT	On Jul 7	2001	to	requests: clone

COMMENT: ON JUL 7, 2001 THE
----- GE

Center: Sanger Ce

Center code: SC

Web site: <http://www.chem.mcgill.ca/~chem222>

Contact: humquery

Id -----

Center project name

```
----- Sub
Assembly program:
```

Assembly program:
Sequencing vector

Chemistry: Dye-templated synthesis of porous polymer networks

Consensus quality

Consensus quality

Consensus quality

```
Insert size: 2415
Insert size: 3393
```

Insert size: 2393
Quality coverage:

quarry coverage:
coverage: 8.51x in

4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

FLSSVTAASGLLYCYELGIISGALLQIKTLITLSCHHEOMVSVSLLIGALLASLTGG
VLIDRYGRRTAIISSCLLGLCSVLILSLSYTVLVIGRIAGVSLISSATLQAYIA
EIAPOHRRLLVLSNELMIVIGILSAYISNFAFANPHGWKMYMFLGVIPLGLQAIAM
YFLPPSPRFLYMKOEGAAKSVLGRALSDATBELAVIKSLKDEYQYSPWDLFRSK
DNMTRIMIGLTVFFVQITGPNILFLYASTLVKSVGFQSNAAASLASTGVGVVFRIS
TIPATLLVDHVGSTFLICIGSSVMAASLVTMGIVNLNTHMFTNICRSHNSINQSLDE
SVIYGPGLNASNNTLRDPHFGIASHSRSSIMPLRNDVDKRGETTASLLNVLSHTE
YQIVTDGDDVPAFLKWLSLASLLVYVAFAFSLGPGMPLVLVSEIFPPGIRGRAMALTS
SMNGINLISLTFLTVTDLGLPWVCFIYTIMSLASLLFVYVMFIPETKGGCSLEBISM
ELAKVNYVKNNICFMSHHOEBLVPRQPKRPQOLLECNKLCGRGQSRQLSPEN*
BASE COUNT 1340 a 943 c 920 g 1318 t
ORIGIN

alignment_scores:

Quality: 932.00 Length: 551
Ratio: 2.596 Gaps: 7
Percent Similarity: 65.154 Percent Identity: 39.201

alignment_block:

US-09-652-292-2 x AB056798

Align seg 1/1 to: AB056798 from: 1 to: 4521

10 LeuCysAlaSerValSerLeuLeuGlyGlyLeuThrPheGlyTyrGluLeu 26
292 CTGTCATCTGTCATCCGCTGCTGAGTGGCCCTCTGGTGGGTTATGAAC 341
26 uAlaValIleSerGlyAlaLeuLeuProLeuGlnLeuAspPheGlyLeu 43
342 TGGGATCATCTCTGGGCTCTCTTCAGATCAAAACCTTATTAAACCTGA 391
43 erCysLeuGluGlnGluPheValIcylSerLeuLeuLeuGlyAlaLeu 59
392 GCTGCCATGACGAGAAATGGTGTGAGCTCCCTCATTTGGAGCCCTC 441
60 LeuAlaSerLeuValGlyPheLeuIleAspCysTyrGlyArgLysG 76
442 CTCGCCCTCACTACCGGAGGGTCTGTATAGACAGGTATGGAAGAGGAC 491
76 nAlaIleLeuGlySerAsnLeuValLeuLeuAlaGlySerLeuThrLeu 93
492 CGCAATCATCTCTGTCATCCCTGCTGCTGGACTCGGAAGCTTAGCTTGA 541
93 lyLeuAlaGlySerLeuAlaThrLeuValLeuGlyArgAlaValGly 109
542 TCCTCAGTTTATCTACACGCTTCTTATAGTGGAGGCATTGCCATAGG 591
110 PheAlaIleSerLeuSerMetAlaCysCysIleTyrValSerGluLeu 126
592 GTCTCCATCTCCCTCCCTCCATTTGGCACTTGTGTTTACATCGCAGAGAT 641
126 uValGlyProArgGlnArgGlyValLeuValSerLeuTyrGluAlaGly 143
642 TGTCTCTCAACACAGAGAGCCCTTCTGTCTCACTGAATGAGCTGATGA 691
143 leThrValGlyIleLeuLeuSerTyrAlaLeuAsnTyrAlaLeuAlaGly 159
692 TTGTCATCGGCATCTCTCTGCTTATATTTCAAAATTACGCATTGGCCAAT 741
160 ThrProTyrGlyTyrPargHisMetPheGlyTyrAlaThrAlaProAlaVa 176
742 GTTTTCCATCGGCTGGAAGTACATGTTGGTCTGTGATTCCCTTGGGAAT 791
176 lLeuGlnSerLeuSerLeuLeuPheLeuProAlaGlyThrAspGluThr 193
792 TTTGGAAGCAATTGCAATGATTATTCTCTCCCAAGC..... 828
193 laThrHisIysAspLeuIleProLeuGlnGly.....GlyGluAlaPro 207
829CCTCGGTTTCTGCTGATGAAAGGACAAAGAGGAGGCTGCTAGC 870
208 LysLeuGlyProGlyArgProArg..... 215

871 AAGTT...CTTGAAGGTTAAGAGCACTCTCAGATCCAATGAGGAAC 917
216TyrSerPheLeuA 220
918 CACTGTGATCAAAATCATCCTTTGAAAGATGAATATCATGATAGTTTGGG 967
220 spLeuPheArgAlaArgAspAsnMetArgGlyArgThrThrValGlyLeu 236
968 ATCTGTTTCGTTCAAAAGACAACATCGGACCCCAATAATGATAGGACTA 1017
237 GlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnValLeuCysTy 253
1018 ACACCTGGTATTTTGTACAAATCACTGGCAGCCCAACATATTTCTCTA 1067
253 rAlaSerThrIlePheSerValGlyPheHisGlyGlySerSerAlaV 270
1068 TGCATCAACTGTTTGAAGTCAGTTGGATTTCAAAGCAATGAGGCAGCA 1117
270 alLeuAlaSerValGlyLeuGlyAlaValIysValAlaAlaThrLeuThr 286
1118 GCCTCGCCTCCACTGGGGTTGGAGTCGTCAAAGTCATTAGCACCATCCCC 1167
287 AlaMetGlyLeuValAspArgAlaGlyArgAlaLeuLeuLeuAlaG 303
1168 GCCACTCTTCTGTAGACCATGTCGGCAGCAAAACATTTCTCTGCATTGG 1217
303 yCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuValSerPheA 320
1218 CTCCTCTGTGATGGCAGCTTCGTTGGTGACCATGGCATCCTAATCTCA 1267
320 laValProMetAsp..... 324
1268 ACATCCACATGAACCTTCACCAATATCTGCAGAGCCCAAAATTTCTATCAAC 1317
325SerGlyProSerCysLeuAlaValPr 333
1318 CAGTCCTTGGATGAGTCTGTGATTATGGACCAGGAAACCTGTCCAGCCAG 1367
333 oAsnAlaThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnA 350
1368 CAACAATACTCTCAGACAGCACCCTTAAAGGAATGCTTCCCATAGCAGAA 1417
350 spSerSerLeuPro.....ProIleProArgThrAsnGluAspGln 363
1418 GCTCACTCATCCCTCAGAAATGATGTGGATAAGAGAGGGGAGCAGCACC 1467
364 ArgGluProIleLeuSerThrAlaLysLysThrLysProHisProArgSe 380
1468 TCAGCATCTTGTCTAAATGCTGTATTAAGCCACACTGAATACCAGATAGT 1517
380 rGlyAspProSerAlaProProArgLeuAlaLeuSerSerAlaLeuProG 397
1518 CACAGACCTCGGAGCAGCTCCA..... 1539
397 lyProProLeuProAlaArgGlyHisAlaLeuLeuArgTyrThrAlaLeu 413
1540GCTTTTGAATGGCTGCTCTTA 1563
414 LeuCysLeuMetValPheValSerAlaPheSerPheGlyPheGlyProVa 430
1564 GCCAGCTTGTGTTATGTGCTGCTTTTCAATTTGTTAGGACCAAT 1613
430 lThrTrpLeuValLeuSerGluIleTyrProValGluIleArgGlyArgA 447
1614 GCCCTGGCTGTGCTCAGCAGATCTTTCCTGGTGGGATCAGAGGAGCAG 1663
447 laPheAlaPheCysAsnSerPheAsnTrpAlaAlaAsnLeuPheIleSer 463
1664 CCATGGCTTTAACTTCTAGCATGAACCTGGGCATCAATCTTCTCATCTCG 1713
464 LeuSerPheLeuAspLeuIleGlyThrIleGlyLeuSerTrpThrPheLe 480
.....

misc_feature

```

22999 ATCTGTTTCTTCAAAAGACAACATCGGACCGCAATAATATGATAGGACTA 22950
237 GlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnValLeuCysty 253
22949 ACATAGTATTTTCTGACAAATCACTGGCGCAACCAACATAATTGTTCTA 22900
253 rAlaSerThrIlePheSerValGlyPheHisGlyGlySerSerAlaVal 270
22899 TGCATCAACTGTTTGAAGTCACTTGGATTTCAAAGCAATGAGCGCAGCTA 22850
270 aLeuAlaSerValGlyLeuGlyAlaValValValValAlaAlaThrLeuThr 286
22849 GCCTCGCTCCATCGGTTGGAGTCTCAAGTCTCATGACCACTCCCT 22800
287 AlaMetGlyLeuValAspArgAlaGlyArgAlaLeuLeuLeuAlaG1 303
22799 GCCACTCTTCTTGTAGACCATGTCGCGCAGCAACCAATTCCTGCAATTGG 22750
303 yCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuValSerPheA 320
22749 CTCTCTGTGATGGCAGCTTCTGTTGGTACCATGGGCATCGTAATCTCA 22700
320 laValProMetAsp..... 324
22699 ACATCCACATGAATCTCACCATATCTGCAGAACCCACAAATTCATCAAC 22650
325 .....SerGlyProSerCysLeuAlaValPr 333
22649 CAGTCTCTGGATGAGTCTGTGATTTATGACCAAGCAACCTGTCAACCAA 22600
333 oAsnAlaThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnA 350
22599 CAACAATACTCTCAGACAGACCACTTCAAAAGGGATTCTTCCCATAGCAGAA 22550
350 spSerSerLeuPro.....ProIleProArgThrAsnGlyAspGln 363
22549 GCTCACTATGCCCCCTCAGAAATGATGTGGATAAGAGGGGAGACCAACC 22500
364 ArgGluProIleLeuSerThrAlaLysThrLysProHisProArgse 380
22499 TCAGCATCTTGTGTAATGCTGATTAAGCCACACTGAATACCAGATAGT 22450
380 rGlyAspProSerAlaProProArgLeuAlaLeuSerSerAlaLeuProG 397
22449 CACAGACCTGGGACGCTCCCA..... 22428
397 lyProProLeuProAlaArgGlyHisAlaLeuLeuArgTrpThrAlaLeu 413
22427 .....GCTTTTGTGAATGGCTGCTCTTA 22404
414 LeuCysLeuMetValPheValSerAlaPheSerPheGlyPheGlyProva 430
22403 GCCAGCTTGTCTTTATGTTGCTGCTTTTCAATTGCTAGGACCAAG 22354
430 lThrTrpLeuValLeu 435
22353 TAAGTACTTATTCTT 22338
seq_name: gb_hhg:AC023293
seq_documentation_block:
LOCUS AC023293 163569 bp DNA HTG 01-MAR-2000
DEFINITION Homo sapiens clone RP11-17F18, WORKING DRAFT SEQUENCE, 17 unordered
pieces.
ACCESSION AC023293
VERSION AC023293.2 GI:7138735
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 163569)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens, clone RP11-17F18
Unpublished
2 (bases 1 to 163569)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campolano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collinmore,A., Cooke,P.,
DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenestor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Holland,J.C., Illey,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Mense,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

TITLE JOURNAL COMMENT

Direct Submission
Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 1, 2000 this sequence version replaced gi:6960435.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3635
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 153600 bases at least Q40
Consensus quality: 158799 bases at least Q30
Consensus quality: 160722 bases at least Q20
Insert size: 176000; agarose-fp
Quality coverage: 4.0 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1467: contig of 1467 bp in length
* 1468 1567: gap of 100 bp
* 1568 3349: contig of 1782 bp in length
* 3350 3449: gap of 100 bp
* 3450 6675: contig of 3226 bp in length
* 6676 6775: gap of 100 bp
* 6776 9516: contig of 2741 bp in length
* 9517 9616: gap of 100 bp
* 9617 13613: contig of 3997 bp in length
* 13614 13713: gap of 100 bp
* 13714 17758: contig of 4045 bp in length
* 17759 17858: gap of 100 bp in length
* 17859 22031: contig of 4173 bp in length
* 22032 22131: gap of 100 bp


```

|||||
55675 GCCTCGCTCCACTGGGTTGGAGTCGTCAAGTCTATTAGCACCATCCCT 55724
287 AlaMetGlyLeuValAspArgAlaGlyArgAlaLeuLeuAlaG1 303
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
55725 GCACCTCTTCTGTAGACCATGTCGGCAGCAAAACATCTCTGCAATGG 55774
303 yCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuValSerPheA 320
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
55775 TCCCTCTGTGATGGCAGCTCGTGTGGTACCATGGGCATCGTAAATCTCA 55824
320 laValProMetAsp..... 324
55825 ACATCCACATGAACCTCACCCATATCTGCAGAACCCACAATCTATCAAC 55874
325 .....SerGlyProSerCysLeuAlaValPr 333
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
55875 CAGTCCTTGGATGAGTCGTGATTTATGCACCGAAGAACCTGTCAACCAA 55924
333 oAsnAlaThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuGlnA 350
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55925 CAACAATACCTCAGACAGACCACTTCAAGGGATTTCTTCCCATAGCAGAA 55974
350 sPserSerLeuPro.....ProIleProArgThrAsnGluAspGln 363
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55975 GCTCACTATGCCCCCTCAGAAATGATGTGGATAAGAGAGGGAGACGACC 56024
364 ArgGluProIleLeuSerThrAlaLysLysThrLysProHisProArgSe 380
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56025 TCAGCATCCTGCTAAATGCTGGATTAAGCCACCACTGAATACCATGACT 56074
380 rGlyAspProSerAlaProProArgLeuAlaLeuSerSerAlaLeuProG 397
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56075 CACAGACCCCTGGGAGCTCCCA..... 56096
397 lyProProLeuProAlaArgGlyHisAlaLeuLeuArgTrpThrAlaLeu 413
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56097 .....GCTTTTGAATGGCTGCTCTTA 56120
414 LeuCysLeuMetValPheValSerAlaPheSerPheGlyPheProVa 430
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56121 GCACGCTGCTGTTTATCTGCTTTTCAATGTTGCTAGGACCAAG 56170
430 lHrTrpLeuValLeu 435
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56171 TAAGTACTTTATCTT 56186
seq_name: gb_bs:BSZ92954
seq_documentation_block: 8189 bp DNA BCT 24-JUN-1998
LOCUS BSZ92954
DEFINITION B-subtilis yws[A,B,C,D,E,F,G] and gerBC genes.
ACCESSION 292954
VERSION 292954.1 GI:1894764
KEYWORDS gerBC gene; ywsc gene; ywta gene; ywtB gene; ywtC gene; ywtD gene; ywTE gene; ywTF gene; ywtG gene.
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillus/Clostridium group:
Bacillus/Staphylococcus group; Bacillus.
1 (bases 1 to 8189)
AUTHORS Prescan,E., Moszer,I., Boursier,L., Cruz,H., De La Fuente,V.,
Hullo,M.F., Lelong,C., Schleich,S., Sekowska,A., Song,B.H.,
Villani,G., Kunst,F., Danchin,A. and Glaser,P.
TITLE The Bacillus subtilis genome from gerBC (311 degrees) to lncR (334
degrees)
JOURNAL Microbiology 143 (Pt 10), 3313-3328 (1997)
MEDLINE 98015417
REFERENCE 2 (bases 1 to 8189)
AUTHORS Lelong,C., Glaser,P., Presecan,E. and Danchin,A.
TITLE Bacillus subtilis gerB downstream sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 8189)

```

Glaser, P.
Direct Submission
Submitted (13-MAR-1997) Philippe Glaser, Regulation de l'Expression
Genetique, Institut Pasteur, 28 Rue du Dr Roux, Paris, 75724,
FRANCE

FEATURES
source 1..8189
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
1..1069
/gene="ywsC"
<1..1069
/gene="ywsc"
/note="product highly similar to Bacillus anthracis CapB
protein"
/codon_start=2
/transl_table=11
/protein_id="CAB07475.1"
/db_xref="GI:1894765"
/db_xref="SPTREMBL:P96736"
/translation="RGKSTVTRLTTLILIEAGYKTVGKTTGTDMRIYWDTPPEKPIK
RKQGNIGPQKEVMRETVBERGANAIVSECMVNPDIYQIIFOEELLANIGVIVNVLE
DHMDVGMTLDEIAEATATIPYNGHLVITDSEYTEFFKOKAKERNKTVIIADNSKIT
DEYLKFEYVWFVFDNANSLALGVAQALGIDETAFKMLNAPPDGAMRLLPLSPSEP
GHEVNGFAANDASSTLNIWRKREIGYTPDDPIIIMCRADRVDRTOQOFANDVLVPIYE
ASELILIGETTEPIVKAEECKIPADKLHDLKLEYKSTDEIMELLKSMHNRVIVGVNI
HGAEPUIEKIHEYKVKQLVS"
1069..1075
/gene="ywta"
join(1069..1075,1084..1533)
/gene="ywta"
1084..1533
/gene="ywta"
/note="product highly similar to Bacillus anthracis CapC
protein"
/codon_start=1
/transl_table=11
/protein_id="CAB07468.1"
/db_xref="GI:1894766"
/db_xref="SPTREMBL:P96737"
/translation="MFGSDLYIALILGLVLLSLIPAEKTVIPAGLVVPGVGLVFNOP
VFILLVLLSLLTYIVKYGLSKFPMILYGRKKFAAMLTIGVLKIAEDFLYPIVPPFI
AEFGIGIIVPLGIANTIQKGLTITTFGSLLLSLGSAFTAMFVTLI"
1538..1544
/gene="ywteB"
1538..2694
/gene="ywteB"
1552..2694
/gene="ywteB"
/note="product highly similar to Bacillus anthracis CapA
protein"
/codon_start=1
/transl_table=11
/protein_id="CAB07469.1"
/db_xref="GI:1894767"
/db_xref="SPTREMBL:P96738"
/translation="MKKLSFHEKLLKLTQKQKKTKNKHVFIAIP1VFVLMFAPFMAG
KATPVKVTKYSDVVDVLSASFVGDIMMGYVKEVTEQKGAOSIYOYVEIPFRASDYVAGN
FENDTVQKNYKQADKEIHLQTNKESVKVLKDMNFTVLNSANNHAMDYGVQMKDITLG
EFAKQLNDIVGAGYSLDAKKKISYQKNGVTIATLGTDVSGKPAKKNTKPGVLP
DPELIPWISSEAKKHADIVVQSHWGGEYDNDNDROQLARMSDAGADIIVGHHPH
VLEPIEVNGTVIPYSLGNFVFDQGTWTRDSALVQHLKKNGTGRPEVTPDITHEAT
2697..2702
/gene="ywteC"
2697..2914
/gene="ywteC"
2709..2876
/gene="ywteC"
/codon_start=1
/transl_table=11
/product="unknown"

terminator	/protein_id="CAB07470.1" /db_xref="GI:1894768" /db_xref="SPTREMBL:P96739" /translation="MKEVKAIWPEVAVAIWFMESAFKFNQDLTDQEKQKIDMEMKNI QQQEEPPVANK"
gene	2880..2914 /gene="ywtC" 3000..4258 /gene="ywtD" 3000..3007 /gene="ywtD" 3017..4258 /gene="ywtD"
RBS	/note="product similar to E. coli NLPc protein and to Listeria species p60-related proteins"
CDS	/codon_start=1 /transl_table=11 /protein_id="CAB07476.1" /db_xref="GI:1894769" /db_xref="SPTREMBL:P96740" /translation="MNTLANKKELLVAVIICFLVPMTKAEIAEADTSSSELIVSEAK NLLGQYKIGETPKGEPDPSGLIQYVFSKADHILPRSNDOYKIGTAVKPNLKPGD LFPKKEGSTVPTHDALYIGDGMVHSTQSGVITINYKSSVWSGTYIGARIRAA DPATADYVVOVEAEKYIGCVYVFGGSPSEGFCGLVOYVFOOALGIYLPRAEQOQ AGEKVPQNIKPQDGVVYFNTYKGTGISHAGIVAGAGGRTIQASRSKVTISVLSDDY KSKMTGIRRFNLIIPKENPTVSEATLYGCEVPYKOGGTPETGRTAGFQVQIVTQKA AGISLPYATSOYNAGTYIKKADLKPGDIVFFOSTSLNFSIYIGNGQVVHVTLNSGVT ITNNTSTYKWKQKAGSIRVQ"
terminator	4259..4293
gene	complement(4292..5170) /gene="ywtE"
CDS	complement(4292..5152) /gene="ywtE"
	/note="product similar to Bacillus subtilis YxeH and Ycse proteins and to E. coli Yida protein"
	/codon_start=1 /transl_table=11 /protein_id="CAB07471.1" /db_xref="GI:1894770" /db_xref="SPTREMBL:P96741"
	/translation="MKCIAIDLDTLLNKESVISAEINRAIKRAVDAGTILVTICTGRA TFDKALLDDDDPIITAANGGTHDGYRLISRTLMDQAGRAIADYLLSKNIVPEVY TDDHLLSPFGGEAKLHAELDIKSNPNQTDLMQAGTQKQGIKPIPHISVFDY GGENYKLLGCFSDMDKLQAKBELKHKKLAQTSQGHIIELIPASSCKGRALTFLA DIVGIEGTQDIYAIGDPSNLMSFEVAGHRIAMENAIDELKEKSTPTVKNSDENGVAIF IDQLSGQYA"
RBS	complement(5163..5170) /gene="ywtE"
gene	5277..5305 /gene="ywtF"
terminator	5277..5305 /gene="ywtF"
gene	/gene="ywtF" complement(5308..6291) /gene="ywtF"
CDS	complement(5308..6276) /gene="ywtF" /note="an in frame TGA stop codon is found at position 5599 to 5601 translation similar to Bacillus subtilis LytR protein"
	/codon_start=1 /pseudo /transl_table=11 /db_xref="PID:ei301651" complement(6285..6291) /gene="ywtF"
RBS	6594..8112 /gene="ywtG"
gene	6594..6601 /gene="ywtG"
RBS	6594..6601 /gene="ywtG"
CDS	6609..7982 /gene="ywtG"
	/note="product highly similar to metabolite transport proteins"

432 rpleuValLeuSerGluIleThrProValGluIleArgGlyArgAlaph 448
7684 GGGTGATGCCTCCCAATTGTTCGGTTCACGTACAGGAATCGGGACC 7733

449 AlaphCysAsnSerPheAsnTrpAlaAlaAsnLeuPheIleSerLeuSe 465
7734 GTGTTTCGACCTTAATGTTCCAGCTTGGGACACTGATTGTTTCATTAA 7783

465 rPheLeuAspLeuLeuGlyThrIleGlyLeuSerTrpThrPheLeuLeu 482
7784 CTATCCAATAATTAATGAAGCGATCGGAATCAGTATTATTTCTCTGATT 7833

482 yrGlyLeuThrAlaValLeuGlyLeuGlyPheIleThrLeuPheValPro 498
7834 ATGCCGGATCGGTATCATGCGCTTCTATTGTTCGGATTAAAGTGACA 7883

499 GluThrLySgInSerLeuAlaGluIleAsgInGlnPheGlnLysAr 515
7884 GAGCAAAGGGAAGAAGCCTTGAAGAATTCAGCAGGATTACGGGACAA 7933

515 gArPheThrLeuSerPheGlyHlsArGlnAsnSerThrGly 529
7934 AAACGGACAGGAGGAGCGCGCGGAACACAGCACTGTCGGA 7976

seq_name: gb_ba:BSUB0019

seq_documentation_block:
BSUB0019 212610 bp DNA BCT 24-JUN-1999
Bacillus subtilis complete genome (section 19 of 21): from 3597091
to 3809700.
ACCESSION Z99122 AL009126
VERSION Z99122.1 GI:2636029
KEYWORDS .
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.

REFERENCE
AUTHORS Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G., Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borcherdt,S., Brion,S., Boursier,L., Brans,A., Braun,N., Brignell,S.C., Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V., Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J., Daniel,R.A., Deniset,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D., Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E., Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A., Galleron,N., Ghim,S.Y., Glasner,P., Goifeau,A., Golightly,E.J., Grandi,G., Gutseppi,G., Guy,B.J., Haga,K., Halech,J., Harwood,C.R., Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hulio,M.F., Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y., Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P., Koningsstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A., Lardinols,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H., Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P., Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M., Ogawa,K., Ogihara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M., Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujić,P., Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M., Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T., Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J., Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B., Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K., Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpestra,P., Tognoni,A., Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A., Viari,A., Wambutt,R., Wedler,E., Wedler,K., Weitzengatter,T., Yoshida,K., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K., Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and Danchin,A.

TITLE The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
JOURNAL Nature 390 (6657), 249-256 (1997)
MEDLINE 98044033

AUTHORS TITLE JOURNAL

Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.

Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES source

Location/Qualifiers
1. .212610
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
complement(10. .3879)
/gene="yvnB"
complement(10. .3879)
/gene="yvnB"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB15521.1"
/db_xref="GI:2636030"
/db_xref="SPTREMBL:O34986"
/translation="MRKVTVIASILLFLSVLSGHHESKAPPVVQQLFQTPHYPL
LENPQIDVKAEMRSNLMLEAQVRDYSQAESAFYKQSNELGKHWMPDPTGAVR
KFLAQPRLIWNSELEYIIVMSNGKQVSETKKLGEGYQADLAHPELLITELAV
DFKNGRADGTEFLFNTDRTIDFYHRYRYPKEGPDSDLIWRDPERILPSGE
TFVWKLKAGHPELTSADFNRYYQTLKEGNAVIDETEGMANTPRVAVISTNGK
DISVAHYRKLARLSSVLYRPLNGTAELNISIGENKPSPTVLOAQPVDOKRKIK
LDKEPVIEDITDRKPVPAESIELRADI RDRSLVKTVAFYRTDENKPFKRIAEKD
RNDLNFHIVYSPELIGKQLEYVVAAGDGNEARTPVKMDIDIKQTSKAGLRMNIE
RDTLSQTOFLKATPEGRADSIKLWDGKKQVPEAMEKEVYFAEDTRKNTLYFNKAVT
MEGKVLKVFDDTNNKRYRYSVLPETLLRKQKQRIIRSGSKVSPDFAENRDEL
VKNARLVSDGTIVRQVRSPKELFIGNQRSNKQFQFNLPDGLFTSQLEWDTLS
KLSEGAHHIQASDGENSVLVNDNSGPHLEPNITEQYKGNLIQADMYDKWSRI
EEAASLDSGISTILPYHTSSDDLPGKHSKVTATDLGKNKTVIERIFKTEREPDOP
EVIDSEADTHRAKLSVRKQPTDAMDVGFRGQYQARNHVLFKHASLTPKPSV
PERPTFTNKLERSAADGKTSTENKELFPHYRFEVTDVPSIDENDLAEVWKGSS
LPGRKVTANNTNENQPVDSFVAKDKKFLTKASVIAADFVRESKMVIYODEIP
PAKMDYTFWMSDQTYTAESYPIFKQTEWIKQNLQKLNKIVYFTGDIYDDSDAIR
QNRKADMSGLDKSGIPYGLAGNHGDKGYSYRAFGYFGSDRDKKPFHYGGSYK
NNRHYDLISSGNDYIMLYMGWGITDEDIAAMNQVLKHPDRMAILAFHEYLVSNG
RSPICEKIFRIVPNPNVNVMLSGHVSAMRKDELDGDKGPDRLVHQLADYOG
GPEGQGVRLRLQFDOANDMVHVSITYPVYKKNYIDTDYGNKDEFSLDLKPRIK
KVETDYFCNVYTNELGKREOVKSGDTAEFRDHLPEQSVYIIIVDSFNKTKS
PIWKFTKKTYRAPQDFRHSNP"
4044. .4517
/gene="yvnA"
4044. .4517
/gene="yvnA"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB15522.1"
/db_xref="GI:2636031"
/db_xref="SPTREMBL:O34692"
/translation="MDNSIHDELFOALQKALRDKRWKQVQIPSIGISSQHHDLK
KDWTLQHLIVSCILHSONVNSFLASRLHISAASKAVHALLKHNITVTKKPGNK
KEIFYLTDSGRKLAALHEQLHEKAEQYKQLFNEFSIDDLKTVTAFFNLWIKYM"
complement(4558. .5775)
/gene="cypX"
complement(4558. .5775)
/gene="cypX"
/codon_start=1
/transl_table=11
/product="cytochrome P450-like enzyme"
/protein_id="CAB15523.1"
/db_xref="GI:2636032"
/db_xref="SPTREMBL:O34926"
/translation="MSQSKLFSVLSQDFONPNPYAFVSOLREEDPVHYEESIDSYFIS
RYHDVRYIQHPDIFTTKSLVERAEPVMRGPVLAOMHGKHSKRIVRYSFIGBALD
HLSPLIKQNAENLAPYLRGKSLDYNDFGKTFACVYMDMLGLDKRHEKISEWHS
VADFITISQSPEARAHSLWCSEQLSQYLLPVIKERRVNPNGSDLSILCTSEYEGMAL

SDKDLTALLLVNLAATSPADKTLALMTYHLLNPNPEQNDVLADRSVPRAIAETLRY
KPPVQLIPRQSQDITVVGGMETIKKDTIVFCMIGANRDPFAEQPDVFNIIREDIGIK
SAFSQAARHLAFAGSGIHNCVGAFAAKNEIEIVANIVLDKMRNIRLEEDFCVAESGLYT
RGPVSLLVAFDGA"
complement(5791. .6537)
/gene="yvmC"
complement(5791. .6537)
/gene="yvmC"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15524.1"
/db_xref="GI:2636033"
/db_xref="SPTREMBL:O34351"
/translation="MTGMVTRRSVHFIAEALTENCREIFERRRHVLVGISPPNSRFS
EDYIYRLIGNAKQFQSVSVLLAGHEAANLLEALGTPRGAERKVRKEVSRNRRFAER
ALVAHGDPKAIHTFSDPIONKAYOLLQROVEHAFFEQPHFRHACLDMKREAIIGRAR
GVSIMMEEVSEDLNLAVEVYIAELPFFIGAPDILEVEETLLATHYRPWKLGKISNHE
FSICHRPNQGLIIVQEAQMLSEKRIITSEG"
6963. .7472
/gene="yvmB"
6963. .7472
/gene="yvmB"
/function="unknown"
/note="alternate gene name: yzhA
similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15525.1"
/db_xref="GI:2636034"
/db_xref="SWISS-PROT:P40762"
/translation="MSDLTKQMIYDIYVRLHLLNEQKANTSLQOFKFAEEDVAEIP
KNTSLHVIDICIGQHEPINNAGIARKMNSKANVTIKLKEEFINSYQLTDNKKKE
VYKTLTRKGRRIEFDLHEKLHKKELAFYQFLDSFOEQKAVLKFLEQLTSTLEAQT
DGTPDPKPVK"
7493. .8726
/gene="yvmA"
7493. .8704
/gene="yvmA"
/function="unknown"
/note="similar to transporter"
/codon_start=1
/transl_table=11
/protein_id="CAB15526.1"
/db_xref="GI:2636035"
/db_xref="SPTREMBL:O34307"
/translation="MAHTKSKAVLILYTVCFSAFFASLSQNIYSPILPIKGSFHVST
AMVNLVSVMIVTAIMQIILGAIIDFKGARIVLTIGILATAASIGCAVTTDFTLFL
IFRMIOAGSAALPLIAATTIGOLFNGERSAMGTQMLLSVAIPAIAVPLGFGGA
AGYEGIFMLAALISIVLLVTNSTITPKDSPTESMOOAKGNVFAHYSKIFNTGTGVAL
TLSFVLFYFAVIVYLPILLTEHYHIDVGIAGLILPLALSTIAGFLFKRIQAKIG
LAHTLFGSNVIAACSTILPVTHTSVSLMALALFGLISMGVPIPLPYSTMTITNFEH
NKGSAIGMFNFIYRTGMAAGPMVSAYLLTMPSAMSFSLLGIGFAALSFCLLPPMFSP
QARTKQKKHHM"
8709. .8726
/gene="yvmA"
complement(8732. .9091)
/gene="yviD"
complement(8732. .9091)
/gene="yviD"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB15527.1"
/db_xref="GI:2636036"
/db_xref="SPTREMBL:O34648"
/translation="MKWAVSILVNALLIIVDGYDSIHISIGGAATIASLILSLN
VLIRKLLIIFLPTVMTKVLSGLFLFVINAITLMTASINGDSFQIDGFGTAIWSVILS
VFHLLIQHIGLLEPLKK"
complement(9093. .9290)
/gene="yviC"
complement(9093. .9290)

terminator

gene
CDS

[illegible]

```

seq_name: gb_ba:BSUB0021
seq_documentation_block:
LOCUS      BSUB0021      215534 bp      DNA      BCT      26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 21 of 21): from 3999281
            to 4214814.
ACCESSION  Z99124 AL009126
VERSION    Z99124.1 GI:2636442
SOURCE     Bacillus subtilis
ORGANISM   Bacillus subtilis
REFERENCE  1 (bases 1 to 215534)
AUTHORS   Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
            Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
            Boriss, R., Boursier, L., Brans, A., Braun, M., Brignelli, S.C.,
            Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
            Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
            Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
            Emerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
            Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Gallizzi, A.,
            Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golligly, E.J.,
            Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
            Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hulio, M.F.,
            Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
            Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetler, P.,
            Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Liu, H.,
            Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A.,
            Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,
            Mizuno, M., Moesti, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
            Ogawa, K., Ogawara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
            Portetle, D., Porwollik, S., Prescott, A.M., Prescan, E., Pujic, P.,
            Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
            Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
            Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
            Sekowska, A., Seter, S.J., Serror, P., Shin, B.S., Soldo, B.,
            Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
            Takeuchi, M., Tanakashi, A., Tanaka, T., Terstra, P., Tognoni, A.,
            Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
            Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T.,
            Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
            Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
            Danchin, A.
            The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
            Nature 390 (6657), 249-256 (1997)
            98044033
2 (bases 1 to 215534)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr.
adanchin@pasteur.fr Phone: +33 (0) 1 45 68 84 41, Fax: +33 (0) 1 45
68 89 48
FEATURES             Location/Qualifiers
     1..215534
        /organism="Bacillus subtilis"
        /strain="168"
        /db_xref="taxon:1423"
        complement(11..820)
        /gene="yxjF"
        complement(11..37)
        /gene="yxjF"
        complement(47..820)
        /gene="yxjF"
        /function="unknown"
        /note="similar to gluconate 5-dehydrogenase"
        /codon_start=1
        /transl_table=11
        /protein_id="CAB15933.1"
gene
terminator
CDS

```


[illegible]


```

133 yValLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeu 150
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30681 GACACTCGCACAATGAACAATTAATGATGTCACCGGATTTTGCTCG 30632

150 eTyrAlaLeuAsnTyrAlaLeuAlaGlyIleThrProTyr...GlyTrpArg 165
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30631 CCTATATGTTAACTACCTGTTTC.....ACACCGTTTGAAGCGTGCGC 30588

166 HisMetPheGlyTrpAlaThrAlaProAlaValLeuGlnSerLeuSerLe 182
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30587 TGGATGGTGGCTGGCTGCATGCTGCTGCTGCTTCTATTAAATCGGCAT 30538

182 uLeuPheLeuPro.....AlaGlyThrAspG 191
: ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30537 CGCCCTTCATGCGGAATCGCAAGATGGCTGCTGAAGCGCGCAGCAAG 30488

191 luThrAla.....ThrHis..... 195
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30487 AAGAAGCGAGCGCATCATGAACATCACTGACCCCAAGACATCGAA 30438

196 LysAspLeuIleProLeuGlnGlyGlyAlaProLysLeuGlyProG1 212
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30437 ATGGAATTGGCAGAGATGAAGCAAGGGAAGCGGAGAA..... 30399

212 yArgProArgTyrSerPheLeuAspLeuPheArgAlaArgAspAsnMetA 229
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30398 .....ANAGAAGCAGACACTTGGCGTGTGAAAGCCCAAG...TGGATTC 30359

229 rgGlyArgThrThrValGlyLeuGlyLeuValLeuPheGlnGlnLeuThr 245
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30358 GCCATGCTTCTGATGGCTGGCGCTTGCATTTCCACGACGCGAGTC 30309

246 GlyGlnProAsnValLeuCysTyrAlaSerThrIlePheSerValG1 262
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30308 GGCATTAAACAGCGTCATCTACTACGCGCCCAACCATTTTACGAAGCGCG 30259

262 yPheHisGlySerSerAlaValLeuAlaSerValGlyLeuGlyAlav 279
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30258 TCTT...GGCACAATCAGCTTGGCGCTGCGCAACAATGGGAATCGGCATCC 30212

279 alLysValAlaAlaThrThrAlaMetGlyLeuValAspArgAlaGly 295
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30211 TCAACGTCATCATGTCATCCTCCATGATCTTAATGACCGTGTGCGC 30162

296 ArgArgAlaLeuLeuAlaGlyCysAlaLeuMetAlaLeuSerValSe 312
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30161 CGCAAAACTTCTGATCTGGGAAGTGTGGCATCACACTGAGCCCTGC 30112

312 rGlyIleGlyLeuValSerPheAlaValProMetAspSerGlyProSerC 329
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30111 TGCTTTA..... 30105

329 ysLeuAlaValProAsnAlaThrGlyGlnThrGlyLeuProGlyAspSer 345
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30104 .....TCA 30102

346 GlyLeuLeuGlnAspSerSerLeuProProIleProArgThrAsnGluAs 362
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30101 GCGTGCTCTCGACGTGGGCTTCCTC..... 30075

362 pGlnArgGluProIleLeuSerThrAlaLysLysThrLysProHisProA 379
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30074 .....GCATCAACACGG..... 30063

379 rgSerGlyAspProSerAlaProProArgLeuAlaLeuSerSerAlaLeu 395
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30063 ..... 30063

396 ProGlyProProLeuProAlaArgGlyHisAlaLeuLeuArgTrpThrAl 412
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30062 .....TGGATGAC 30055

```

```

412 aLeuLeuCysLeuMetValPheValSerAlaPheSerPheGlyPheGlyP 429
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30054 AGTAGTCTTTTCGGTGCTATATCGTCTTTTACACGCGACATGGGTC 30005

429 roValThrTrpLeuValLeuSerGluIleTyrProValGluIleArgGly 445
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30004 CGGTAGTGTGGTGCTCATGCTGAGCTTCCTCCGTCACAAAGCAGCGGA 29955

446 ArgAlaPheAlaPheCysAsnSerPheAsnTrpAlaAlaAsnLeuPheI 462
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
29954 GCCGCGACAGGTTTCACGACTTTGGTGTATCGCGCGCTAACCTGATCGT 29905

462 eSerLeuSerPheLeuAspLeuIleGlyThrIleGlyLeuSerTrpThrP 479
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
29904 GTCGCTGTCTCCGCTGATGCTAGGCCGATGGCATCGCATGGGTCT 29855

479 heLeuLeuTyrGlyLeuThrAlaValLeuGlyLeuGlyPheIleTyrLeu 495
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
29854 TTATGGTCTTCTCAGTCATTTGCTCTCTCTCATTTTCTTCGCGCTTCTAC 29805

496 PheValProGluThrLysGlyGlnSerLeuAlaGluIleAspGlnGlnPh 512
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
29804 ATGTACCGGAACCAAGGAAGCTTGGAGAAATTGAAGCAAGCCT 29755

512 eGlnLysArg 515
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
29754 GAAAAACGG 29745

seq_name: gb_ba:AF045552
seq_documentation_block:
LOCUS AF045552 5957 bp DNA BCT 12-DEC-1998
DEFINITION Lactobacillus brevis xyl operon and xylose isomerase (xylA),
xylulokinase (xylB), and D-xylose proton-symporter (xylT) genes,
complete cds.
ACCESSION AF045552
VERSION AF045552.1 GI:2895853
KEYWORDS Lactobacillus brevis.
SOURCE Lactobacillus brevis.
ORGANISM Lactobacillus brevis.
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
Lactobacillus.
REFERENCE 1 (bases 1 to 5957)
AUTHORS Chailou, S., Bor, Y.-C., Batt, C.A., Postma, P.W. and Pouwels, P.H.
TITLE Molecular cloning and functional expression in Lactobacillus
plantarum 80 of xylT, encoding the D-xylose-H+ symporter of
Lactobacillus brevis
JOURNAL Appl. Environ. Microbiol. 64 (12), 4720-4728 (1998)
MEDLINE 99054905
REFERENCE 2 (bases 1 to 5957)
AUTHORS Bor, Y.-C. and Batt, C.A.
TITLE The D-xylose operon of Lactobacillus brevis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5957)
AUTHORS Bor, Y.-C. and Batt, C.A.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1998) Food Science, Cornell University, 312
Stocking Hall, Ithaca, NY 14853, USA
FEATURES
source 1..5957
/organism="Lactobacillus brevis"
/db_xref="taxon:1580"
1230..5880
/gene="xyl operon"
1230..1253
/gene="xyl operon"
/note="xyl operator"
1271..2688
/gene="xylA"
1271..2620
/gene="xylA"
/codon_start=1
/transl_table=11

```

```

/product="xylose isomerase"
/db_xref="GI:2895125.1"
/translacion="MTEYWKGVDRKIQYVGHQDKKSGLFQYVNPPEEIMGKKMKDWL
RFAYVWHITFOORLVDPDGTAGRPYDKYTDPMDLALAKYDAAFERYOKLGVDFLCF
HORDLAPDGLTRETANLKVDPKIVYOKTSCMKVLKNTSNMTPNPRFVEGATSP
YADVPAYSAAGLKHSLGKRGVSENYVFGREGYESLWNTNKKQEGEHAAKLPHMA
KDYANEIGDQMLLEPAPKPTTHQYDFDAATTFAPMKYDLDKDFKLNLEGNHL
AGHTYQHEIRVARAGGLDNLNOGDKLIGDIDEXPSNLYETTAAMYEVENGISG
PRGLNFDAKRRFAFEPEDLGHIVGMDSFAGLRVAAMKODGFLDNLKADRYSS
YKSGVGADIESGKADLKLEAYADKPOSELIAATSHDHLSEIKDTHNYIIDLK"
2661..2688
terminator
/gene="xylA"
/ncote="potential Rho-independent transcription terminator"
2724..4287
/gene="xylB"
2724..4232
CDS
/gene="xylB"
/codon_start=1
/translacion="MGKVVGLDGLGTSAAKVSALDHSQIVAQESFDYDLIQKOPGYN
EQNPEDWYSGTVAIVRLILNDHLDASNIELSGSQMHGLVLDENKKVLRPAILLN
DYRTPQREEEALGLDEFTVHTRNPLEGFTLTKLVKQNEPDIMAKAFYVLPKD
YRVTMTGNLAMDYSDATGVLLDVAKGNSQKCAALDIPMSMCPPLIKYIDLATV
TPAVGESGLTDTDFVGGGAADNAGAVGAGILHNPVLSSTGTVGVVLYKVEDNADVN
YHGVLOFEDHAIPOKFSYMGVTLAAGYSFTWPKTFAPADEFTDVAASAKSTYVGANG
LLYTPYIVGERAPYADADIRGFTGVGTHQYDFVRAVLEGLIFSFRLDPIVEENG
GDFTVYVSGGAKSLQIQAIDFNKRVKVSLENEQPGMGCAAMIAATGLWDFDSLQ
DCAETVHFHGKAYEPNPONVKYKEMHAIYKQVYQOQTKTISEQMLLDYRRAL"
4288..4287
terminator
/gene="xylB"
/ncote="potential Rho-independent transcription terminator"
4344..4367
misc_feature
/gene="xyl operon"
/ncote="xyl operator"
4454..5880
gene
4454..5827
CDS
/gene="xylT"
/gene="xylT"
/codon_start=1
/translacion="MRKVSTGVTFVFFGALGGLLFCYDTGVISGAILFLQKONLGSWO
QGWVAVALLGAILGAAIIGPSSDRPGRKLLLSAILFFVGLGASAPSPREWTLIIS
RIILGMVGAASALIPYLAELAPSDRGKTVSSLFQJLVMTGILLIAYITNYSFGFYT
GWRMLGPAALIPAAALLFLGLLIPESRFLVSGHLEARHVLDTMNNKHQDVAVNKEI
NDIQESAKIVSGGSELSFGKMRPRLIIGIGLAIPOQVVGNCNTVLYAPTFTDVGFG
VSAALLAHIGTGFVNIYATAVAIMDKIRKIVNICAVGNGISLFSVMSIKMPSFGG
SOTAAISLIVALTYYIAPFSATWGPVMVMVIGEVFPLNIRGLNSFSASVINWNTANMIV
SLTPFSLLDFGCTGSLFYGILCFASIFWVQKVFETRNRSLIEDIATLRAKYGEDA
AELSTTK"
5842..5880
terminator
/gene="xylT"
/ncote="potential Rho-independent transcription terminator"
BASE COUNT 1600 a 1244 c 1444 g 1669 t
ORIGIN

alignment_scores:
  Quality: 571.00 Length: 515
  Ratio: 1.936 Gaps: 7
  Percent Similarity: 57.282 Percent Identity: 27.379

alignment_block:
  us-09-652-292-2 x AF045552
  Align seg 1/1 to: AF045552 from: 1 to: 5957

```

```

17 LeuGlyGlyLeuThrPheGlyTyrGluLeuAlaValIleSerGlyAlaLe 33
4496 CTGGGTGGGTATCTTTTGGGTATGACACTGGGGTTATTTCGGGAGCTAT 4545
33 uLeuProLeuGlnLeuAspPheGlyLeuSerCysLeuGluGlnGluPheL 50
4546 TCCTTTTATTCAGAACACAGATGACCTTAGGATCTTGGCAACAAGCTGGG 4595
50 euValGlySerLeuLeuGlyAlaLeuLeuAlaSerLeuValGlyGly 66
4596 TCCTTAGTGGGTTTGTAGGGGCCATTTTAGGGGCTGCCATCATTTGGG 4645
67 PheLeuIleAspCysTyrGlyArgLysGlnAlaIleLeuGlySerAsnLe 83
4646 CCTTCGCTGACCGATTTGGTCTGGCGGAATTTATTACTCTTCTGCTAT 4695
83 uValLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAlaL 100
4696 TATTTTCTTCTGGTGGGCGGTAGGTTCCGGCTTTTCCACAGAGTTTGGGA 4745
100 rLeuValLeuGlyArgAlaValIleGlyPheAlaIleSerLeuSerSer 116
4746 CGTTAATTTATCTCCGAATTTATCTGGGATGGCGTGGGCTGCTTCA 4795
117 MetAlaCysCysIleTyrValSerGluLeuValGlyProArgGlnArgG 133
4796 GCATTGATCCGACGATTTTGGCTGAATTTGGCACCGCTCTGATAGCGAGG 4845
133 yValLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeu 150
4846 AACGGTTTCCAGTCTATTTCCGTTAATGGTTATGACTGGGATCTCTTGG 4895
150 erTyrAlaLeuAsnTyrAlaLeuAlaGlyThrProTyrGlyTyrArgHis 166
4896 CCTACATCAGCAACTACAGTTTCTCAGGTTTCTATACCGGCTGGCGTGG 4945
167 MetPheGlyTrpAlaThrAlaProAlaValLeuGlnSerLeuLeuLe 183
4946 ATGTTAGGTTTCGGCGCAATTCCTGCTGCTTCTGCTTCTTGGTGGGCT 4995
183 uPheLeuProAlaGlyThr..... 189
4996 AATTCCTCCCGAGAGTCCCGGCTTCTTGGTTAAGAGTGGGCATCTGGATG 5045
190 .....AspGluThrAlaThr 194
5046 AAGCTCGGCACGTGCTAGACACCATGAAACAGCATGACCAAGTGGCGGTC 5095
195 HisLysAspLeuLeuProLeuGlnGlyGlyGluAlaProLysLeuGlyPr 211
5096 AATAAAGAAATTAATGATATTCAA.....GAATCCGCTAAAAATTTGTAG 5139
211 oGlyArgProArgTyrSerPheLeuAspLeuPheArgAlaArgAspAsn 228
5140 CGGT.....GGTTGGTCAGAGTTTATTTGGTAAGATG.....G 5171
228 etArgGlyArgThrValIleGlyLeuValLeuPheGlnGlnLeu 244
5172 TTGCGCTTCATTCATGATCGGGATGTTTACCCATTTTCCACACAGGTG 5221
245 ThrGlyClnProAsnValLeuCysTyrAlaSerThrIlePheSerSerVa 261
5222 ATGGGCTGTAAACGGGTACTGTACTACGCCACCAACAATCTTACGGATGT 5271
261 lGlyPheHisGlyCysSerAlaValLeuAlaSerValGlyLeuGly 278
5272 TGGCTTC...GGAGTTTCACCGCACTGTTAGCGCATATTGGGAATTGGGA 5318
278 laValLysValAlaIleThrLeuThrAlaMetGlyLeuValAspArgAla 294
5319 TCCTTAACGTGATTGTGACCGCGGATTCGGGTAGCCATCATGGATAAGATC 5368
295 GlyArgAlaLeuLeuLeuAlaGlyCysAlaLeuMetAlaLeuSerVa 311

```


[illegible]

THIS PAGE BLANK (USPTO)